**Mitigating user frustration through adaptive feedback based on human automation etiquette strategies**


by

**Euijung Yang**


A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY


Major: Industrial Engineering


Program of Study Committee:
Michael Dorneich, Major Professor
Stephen Gilbert
Caroline Hayes
Jonathan Kelly
Richard Stone


Iowa State University

Ames, Iowa

2016

TABLE OF CONTENTS

# LIST OF FIGURES

## LIST OF TABLES

# ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this dissertation.

First, I would like to thank my advisor, Dr. Michael Dorneich for his incredible guidance, patience, and support throughout the course of this research. His insights and words of encouragement have always inspired me and renewed my hopes for completing my graduate education. I would also like to thank my committee members, Dr. Stephen Gilbert, Dr. Richard Stone, Dr. Caroline Hayes, and Dr. Jonathan Kelly, for their efforts and contributions to this study.

I would also like to thank my mom Sookhee Oh, my dad Sungkyung Yang, my sister Soojung Yang, and my brother Yeonsoo Yang whose advice, attitudes, and actions have been such a great positive influence on my life and my work.

I would additionally like to thank my fiancé David Montealegre who accompanied me along this journey. He gave me precious advice on my life and constantly encouraged me to overcome any difficulties and challenges I encountered.

Finally, I would like to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience. I want to also offer my appreciation to those who were willing to participate in my experiments and surveys, without whom, this dissertation would not have been possible.

ABSTRACT

The objective of this study is to investigate the effects of feedback and user frustration in human-computer interaction (HCI) and examine how to mitigate user frustration through feedback based on human-automation etiquette strategies. User frustration in HCI indicates a negative feeling that occurs when efforts to achieve a goal are impeded. User frustration impacts not only the communication with the computer itself, but also productivity, learning, and cognitive workload. Affect-aware systems have been studied to recognize user emotions and respond in different ways. Affect-aware systems need to be adaptive systems that change their behavior depending on users' emotions. Adaptive systems have four categories of adaptations. Previous research has focused on primarily function allocation and to a lesser extent information content and task scheduling. However, the fourth approach, changing the interaction styles is the least explored because of the interplay of human factors considerations. Three interlinked studies were conducted to investigate the consequences of user frustration and explore mitigation techniques. Study 1 showed that delayed feedback from the system led to higher user frustration, anger, cognitive workload, and physiological arousal. In addition, delayed feedback decreased task performance and system usability in a human-robot interaction (HRI) context. Study 2 evaluated a possible approach of mitigating user frustration by applying human-human etiquette strategies in a tutoring context. The results of Study 2 showed that changing etiquette strategies led to changes in performance, motivation, confidence, and satisfaction. The most effective etiquette strategies changed when users were frustrated. Based on these results, an adaptive tutoring system prototype was developed and evaluated in Study 3. By utilizing a rule set derived from Study 2, the tutor was able to use different automation etiquette strategies to target and

improve motivation, confidence, satisfaction, and performance using different strategies, under different levels of user frustration. This work establishes that changing the interaction style alone of a computer tutor can affect a user's motivation, confidence, satisfaction, and performance. Furthermore, the beneficial effect of changing etiquette strategies is greater when users are frustrated. This work provides a basis for future work to develop affect-aware adaptive systems to mitigate user frustration.

CHAPTER I: INTRODUCTION

The purpose of this research is to investigate how to mitigate user frustration and improve task performance based on changes to the interaction style between the user and the system in the context of human-computer interaction (HCI). User frustration plays a role in many aspects of HCI and studies have looked into the consequences of frustration in various fields. For instance, user frustration has been diminished by providing a function for users to express their feelings (Klein, Moon, & Picard, 2002), mirroring users' emotions to show the empathy (Woolf, Burleson, Arroyo, Dragon, Cooper, & Picard, 2009), and supporting users with encouragement to induce positive emotions (Graesser, Chipman, Haynes, & Olney, 2005; Azevedo et al., 2009). If a system has methods for detecting a variety of user emotions, it could vary the style in which it delivers feedback to users. For example, an intelligent tutoring system MetaTutor (an affective learning companion) changed the level of encouragement when it detected the users' negative emotions (Azevedo et al., 2009; VanLehn et al., 2014). Furthermore, once a computer system has the ability to understand and include user emotions as factors in human-machine interaction, the communication between users and computer systems could be more realistic, advanced, and sophisticated. With an understanding of how the interaction style affects human's emotions and performance, it is possible that a computer system could effectively mimic a human's ability to change its interaction style in reaction to a human user's emotions. This work provides the foundation for the design of adaptive systems that adapt the interaction styles of feedback.

User Frustration in Human-Computer Interaction

Human emotion plays a role in many aspects of human-computer interaction (HCI). Emotion is a key factor in communication since it can drive the way humans convey information (Ferdig and Mishra, 2004). Previous studies have found that both positive emotions (e.g. motivation) and negative emotions (e.g. frustration) are key components of learning (Woolf, Burleson, Arroyo, Dragon, Cooper, & Picard, 2009; Fisher & Noble, 2009). Negative emotions, especially frustration, are significant factors which lead to lower task performance (Waterhouse & Child, 1953; Solkoff, Todd, & Screven, 1964; Spector, 1975; Klein, Moon, & Picard, 2002; Powers, Rauh, Henning, Buck, & West, 2011), longer time for decision making (Toda, 1980; Bechara, 2004; Lerner, Li, Valdesolo, & Kassam, 2015), and decreasing learning (Graesser, Chipman, Haynes, & Olney, 2005; Fisher & Noble, 2009). These studies looked into the consequences of frustration on various fields. However, mitigation of frustration through system changes has been less explored, given the complexity of the interplay between frustration and HCI.

Frustration, defined as an emotional state in which obstacles impair the progress towards achieving a goal (Lawson, 1965), is a complex emotion related to anger and disappointment; aggression is one of its consequences (Dollard, Miller, Doob, Mowrer, & Sears, 1939). Frustration has been shown to reduce the quality of ongoing performance by eliciting responses that interfered with the completion of a given task (Waterhouse and Child, 1953).

Despite the ongoing technological innovations, frustration remains a common problem for users of computers, personal handheld devices, automated systems, or other computer systems (Bessiere, Newhagen, Robinson, & Shneiderman, 2006; Lazar, Jones, & Shneiderman, 2006). As such, frustration has become of significant interest in the context of HCI. Frustration

has been shown to be both frequent and damaging to productivity. Frustration stemming from the use of computers causes users to waste an average of 42-43% of their time (Lazar, Jones, & Shneiderman, 2006). Previous work found that task performance is influenced by the level of frustration. For example, a higher level of frustration led to lower performance score on digit-symbol substitution test (Hokanson & Burgess, 1964). ). Frustration has led to lower user satisfaction, lower motivation, and drove the users to seek alternative systems (Hoxmeier & DiCesare, 2000; Lazar, Jones, & Shneiderman, 2006). In learning, higher frustration caused slower response times (Chen, Gross, Stanton, & Amsel, 1981) and delayed content acquisition (Amsel, 1992). Frustration also reduced the motivation of students (Weiner, 1985), and lead to a lack of confidence of students in computer science (Hansen & Eddy, 2007).

User frustration is categorized into *integral frustration* (i.e., where the cause of the emotion is from the task itself) or *incidental frustration* (i.e., where the emotion is caused by something outside the task) (Bodenhausen, 1993; Jeon, Walker, & Yim, 2014). Integral frustration's causes are unachievable difficulty levels of the task or lack of resources to solve the task. In these cases, changing the task difficulty levels or providing more resources could be effective. On the other hand, incidental frustration's causes are users' previous experiences or lack of confidence (Brown, 1954; Bodenhausen, 1993; Jeon, Walker, & Yim, 2014). In these cases, understanding user frustration and adapting the behavior of the system could be used to improve the interactions between the human and the system, and potentially decrease user frustration and improve task performance. Once users experience frustration, a computer system needs to detect the emotion and change its behavior to diminish it. Computer systems that adapt its behavior could mitigate user frustration by changing the way it communicates with users, much in the same way a human tutor would change his or her feedback when students become

frustrated. The ways of interpreting the characteristics of user frustration and its influences need to be scrutinized in order to investigate the methods to mitigate frustration.

In human-human interaction, people interact differently when they detect the emotional states of others (Ekman, 1970; Picard et al., 2004). Likewise, computers could potentially react differently when they detect user frustration. By changing the interaction style of the computers, users' feeling could also be changed (Woolf, Burleson, Arroyo, Dragon, Cooper, & Picard, 2009). Some initial studies have been conducted to explore the effect of different interaction styles and the concept of etiquette to see how different feedback could potentially provide better communication in human-human tutoring (Pearson, Kreuz, Zwaan, & Graesser, 1995), situation awareness (Wu, Miller, Funk, & Vikili, 2010), and reliability of the system (Parasuraman & Miller, 2004). For instance, Pearson et al. (1995) used etiquette strategies to understand what interaction styles human tutors use when teaching students. Observations from conversation examples between human tutors and students showed that positive politeness was used to encourage the students when they struggled to solve problems. However, tutors' feedback could also lead to negative impressions for students even though it was not the intention of the tutor (Pearson, Kreuz, Zwaan, & Graesser, 1995). These previous works provided a motivation to examine the interaction styles with etiquette in depth in HCI.

Feedback Approaches to Mitigate Frustration

Observing the ways to communicate from human-human interaction can provide inspiration to design for HCI. When humans interact with each other, their social behaviors are governed by expectations based on conventional norms (Brown and Levinson, 1978; Mills, 2003). These expectations for human-human interaction would be from speaker to hearer and

vice versa. It is possible to have these expectations when people interact with computers in terms of HCI. Etiquette is defined as a code that indicates conventional requirements for social behavior, and the word itself started to be used around 1750 (Oxford English Dictionary, 2015). Interactions between people with inappropriate etiquette may be confusing, unproductive, or even dangerous since people who share the same model of etiquette expect the same level of social behaviors from each other (Wu, Miller, Funk, & Vikili, 2010). The concept of automation etiquette has been introduced to be studied as the application of human-human etiquette conventions to HCI. (Miller & Funk, 2001; Miller et al., 2004; 2005; 2006; 2007).

One approach to designing responsive interactions between human and computer is adaptive systems. An adaptive system can adjust their feedback to users by tracking the condition of their users (Feigh, Dorneich, & Hayes, 2012). Adaptive systems have four categories of adaptations: 1) adjusting the allocation of function between the human and the automation, 2) adaptive the information displayed to the user; 3) changing the tasks of the user by directing their attention, and 4) changing the interaction style between the human and the system. Of these four approaches, changing the interaction styles is the least explored because of the interplay of human factors considerations. For instance, while changing the interaction style is a typical human trait when faced with certain situations, in HCI changing the way information is delivered to users may be a violation of consistency in the human factors criteria (Feigh, Dorneich, & Hayes, 2012). Although adjusting system feedback might violate system's consistency, it does not mean that consistent behavior of system always provides the best results. Varying the ways to render the information to users by reacting to their emotions could produce better performance since the user emotion is one of the factors that drives overall performance in HCI (Klein, Moon, & Picard, 2002).

A system that has the ability to understand and include user emotions as factors in HCI are called affect-aware systems. Affect is used to indicate the experience of feeling or emotion in psychology (Martin, Hogg, & Abrams, 2010); it is also considered as an important factor in personal and social life (Izard, Kagan, & Zajonc, 1984). Affect-aware systems are any kind of systems that contains an ability to consider a user's emotions as an element of the system (D'Mello et al., 2008; Woolf et al., 2009). These are implemented to deliver different types of feedback including encouragement, empathy, and mirroring emotions of the users based on their emotional states (Picard et al., 2004; Picard, 2006; Woolf et al., 2009; Calvo & D'Mello, 2012). Affect-aware systems have the ability to adjust its behavior to consider the users' emotions beyond their performance when they experience negative emotions (Kort, Reilly, & Picard, 2001). If the computers could be more attuned and affect-aware, they may be able to provide the appropriate responses in stressful situations where human emotion is impacting their ability to function.

Feedback and Learning

Feedback is an indispensable element of education, it facilitates growth in a student's potential at different stages in learning. It also enables the students to become aware of their strengths, weaknesses, and which actions they need to employ to improve achievement (Hattie & Timperley, 2007; Norcini, 2010). Studies developed ways to provide effective feedback to enhance students' learning. For example, Wiggins (2012) suggested seven keys to increase feedback effectiveness: goal-referenced, tangible and transparent, actionable, student-friendly, timely, ongoing, and consistent. In a similar manner, methods to provide effective feedback were

developed: specific, realistic, concentrated on student behavior, balanced content, and continued support (Dempsey, 1993; London, 2003; McGill & Brockbank, 2003).

Analyzing processes of learning and problem solving have been advantageous to foster productive feedback. A tutor's awareness of a student's problem-solving process is crucial because the tutor frequently interferes in the student's problem-solving to provide timely feedback. Guiding students through specified problem-solving steps allows the tutors to teach students the ways to approach and solve a given problem (Gordon & Bruch, 1974; Dickman & Gordon, 1985). While analyzing learning processes, students' perceptions are crucial components (Keller, 1987; Keller & Kopp, 1987). The ARCS model was developed to investigate effective ways of interpreting students' perceptions during learning. The ARCS model is a systematic approach, which has four elements for encouraging and sustaining students' motivation in the learning processes: attention, relevance, confidence, and satisfaction (Keller, 1987; Keller & Suzuki, 1988; Keller, 2009).

Pedagogical studies explored the effectiveness of the ARCS model. For instance, the impact of the ARCS model's ability to overcome the lack of motivation was examined by applying it to distance learning students. The ARCS model approach increased students' attention during instruction, established relevant feedback to their needs, generated a positive impression for their confidence, and provided satisfying experiences by emphasizing their achievements. After providing feedback based on the ARCS model, student's motivation to learn was improved (Malik, 2014). For employees who attended staff development classes, their motivation to learn was increased by receiving techniques based on the ARCS model, such as various supporting materials, stories relevant to the learning contents, motivational messages, and compliments on their learning attitudes (Visser & Keller, 1990). Higher levels of motivation,

confidence, perceived satisfaction, and overall performance lead to higher rates of student engagement. These results demonstrated that those four factors enable tutors to provide encouragement and promote students' effective learning (Mohammad & Job, 2012). Similarly, an online feedback system in virtual environments leads to higher levels of motivation, satisfaction, and performance when compared to no-feedback (Geister, Konradt, & Hertel, 2006).

## Objectives and Research Questions

This research investigates how to mitigate user frustration and improve task performance in the context of human-computer interaction (HCI). Understanding the effects of etiquette strategies on users' performances, preferences, and motivations can contribute to a design of an effective HCI system. Furthermore, the selection of proper etiquette strategies for a given situation of the system could mitigate student frustration. Investigation into the effects and mitigation of human emotion in HCI will be explored through three linked studies, which attempt to answer the following three research questions:

- How does feedback influence user emotions, cognitive workload, task performance, and physiological response?
- Does changing the interaction style of feedback by using etiquette strategies mitigate user frustration?
- Does an adaptive system that employs etiquette strategies mitigate user frustration and improve motivation, confidence, satisfaction, and performance?

Three inter-related studies investigated this set of research questions, as illustrated in Figure 1. These studies employ different types of feedback to investigate the effect of user

frustration, examine its consequences, and explore the application of automation etiquette

principles as mitigation techniques.

**Study 1**

RQ1. How does feedback influence user emotions, cognitive workload, performance, and physiological response?

**Manipulate**
*System Feedback*

**Measure**
*Cognitive Workload Physiological Arousal Performance*

**Measure**
*Frustration*
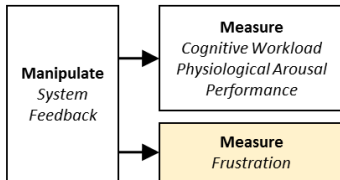
**Study 2**

RQ2. Does changing the interaction style of feedback by using etiquette strategies mitigate user frustration?

**Manipulate**
*Frustration*

**Etiquette Strategies**
*Bald, Positive Politeness, Negative Politeness, Off-Record, No Feedback*

**Provide**
*Feedback*

**Measure**
*Motivation Confidence Satisfaction Performance*

**Assess**
*Impact of Motivation Confidence Satisfaction Performance*

**Rules**

**Study 3**

*Rules*

RQ3. Does an adaptive system that employs etiquette strategies mitigate user frustration and improve motivation, confidence, satisfaction, and performance?

**Target**
*Which Factor?*

**Select**
*Etiquette Strategies*

**Observe**
*Problem Solving*

**Need**
*Feedback?*

Yes

No

**Provide**
*Feedback*

**Start**
*Tutoring Session*

**Measure**
*Motivation, Confidence, Satisfaction, Performance*

**Manipulate**
*Frustration*

**Finish**
*Problem Solving?*

No

Yes
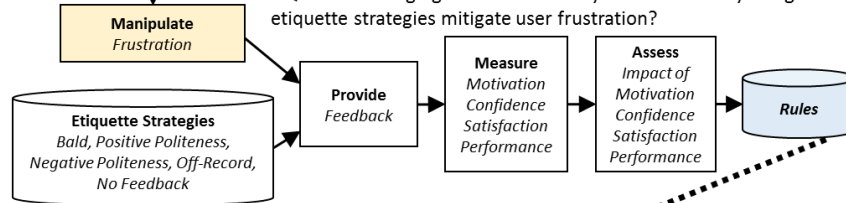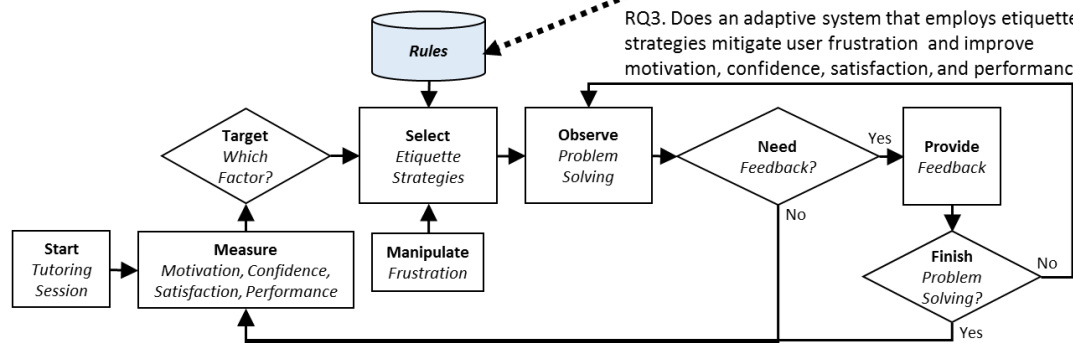
**Figure 1. The overall vision of study.**

Study 1 investigated how feedback in human-robot interaction (HRI) system impacted

user emotional responses, cognitive workload, task performance, physiological arousal, and

usability. This study addressed the first research question by finding that feedback significantly

influences physiological arousal, emotional states (frustration, anger), cognitive workload, and

task performance. Human-robotic interaction was chosen as a domain because it provided a real-

time, interactive testbed to explore the relationships between frustration, feedback, and user

states. After Study 1, there is a transition from HRI to learning domain to investigate the

relationships between frustration, feedback and etiquette strategies. Study 2 explored whether human-human etiquette strategies could be applied to mitigate user frustration while increasing performance, motivation, confidence, and satisfaction in tutoring. The results established etiquette strategies were a viable approach to mitigate frustration and established a set of rules on how etiquette strategies could be used to support student learning. This study addressed the second research question by demonstrating that changing the interaction style of feedback based on etiquette strategies could mitigate the effects of user frustration. Study 3 evaluated an adaptive tutoring system prototype that was able to respond to user frustration and fluctuations in motivation, confidence, satisfaction, and performance by changing etiquette strategies dynamically during tutoring. The results of this study addressed the third research question by demonstrating the effectiveness of an adaptive system based on varying etiquette strategies.

Dissertation Organization

The remainder of this dissertation is organized as follows. Each of the three studies mentioned above has been submitted as a journal paper, which is reproduced here. Chapter II contains Study 1 which investigated the emotional, cognitive, physiological, and performance effects of time delay in robotic teleoperation. In Chapter III contains Study 2 where human-automation etiquette strategies were proposed to mitigate frustration and enhance learning. Chapter IV contains Study 3 which tested the effectiveness of dynamically adapting the automation etiquette strategies during math tutoring, as realized in an adaptive tutoring system prototype. Conclusions, contributions, and future work are discussed in Chapter V.

CHAPTER II: THE EMOTIONAL, COGNITIVE, PHYSIOLOGICAL, AND PERFORMANCE

EFFECTS OF TIME DELAY IN ROBOTIC TELEOPERATION

Euijung Yang and Michael C. Dorneich

Abstract

The effects of time delay were investigated to understand the cognitive and physical consequences of gaps between an input from an operator and the corresponding feedback response from the system. Time delay has been shown to disrupt task performance in various areas including psychology and telerobotics. Previous research in multiple domains has focused on the performance effects of time delay and overcoming technological limitations that cause time delay. However, robotics researchers have yet to study the effects of time delay on specific operator emotions, usability, and physiological activation in teleoperations. This study investigates the influence of time delay not only on task performance, but also operator emotions, physiological arousal, cognitive workload, and usability in teleoperation. Time delay was manipulated by introducing lag into the system feedback. Participants were asked to navigate a remote-control robot vehicle through different mazes in a remote location and simultaneously identify targets. Operator frustration, anger, and workload increased while usability and task performance decreased when feedback lag was introduced to a robotic navigation task. In addition, higher electrodermal activity occurred during time delay conditions. A better understanding of the emotional experiences of human operators and the corresponding

physiological signals is of crucial importance to designing affect-aware robotic systems that have the ability to appropriately respond to operator emotional states.

## Introduction

Despite technological advancements made in the responsiveness of robotic systems, time delay remains a limitation in teleoperation. In this paper, time delay is defined as the gap between an input from an operator and the corresponding feedback response from the system. Telerobotics is a combination of teleoperation and telepresence. Teleoperation is defined as the remote control of semi-autonomous robots, which uses technologies such as wireless networks (i.e., radio control devices, Wi-Fi, Bluetooth, infrared controllers) or wired connections (Sheridan, 1992; Satava & Simon, 1993). Remote-controlled robots enable activities in environments that might otherwise be inaccessible or hazardous to humans. Practical applications include areas such as bomb disposal (Drascic, Milgram, & Grodski, 1989), radioactive environment maintenance (Draper, 1993), surgery (Casals, 1998), mine detection (Nonami, Shimoi, Huang, Komizo, & Uchida, 2000), and subsea manipulation (Ridao, Carreras, Hernandez, & Palomeras, 2007).

There are three main causes of time delay in teleoperation: distance (e.g., space operations), processing time of the computer (e.g., computer communication channels), and environment near the robot (e.g., rugged surface under the robot) (Arcara & Melchiorri, 2002; Behnke, Egorova, Gloye, Rojas, & Simon, 2004). Both transmission time for interactions between the controller and the worksite, and the processing time for interpreting signals are common causes of time delay in teleoperation (Held, & Durlach, 1991; Fabrlzio, Lee, Chan, Stoianovici, Jarrett, Yang, & Kavoussi, 2000). Time delay frequently arises during the

transmission of information between a controller and a robot (Prewett, Johnson, Saboe, Elliott, & Coovert, 2010; Owen-Hill, Suárez-Ruiz, Ferre, & Aracil, 2014), and it is a relatively common problem that occurs when transmitting signals across long distances (Corde Lane et al., 2002).

Time delay is disruptive to task performance in teleoperation. Delayed responses from system-induced lag can lead to increased stress, aversive behavior, impatience, and irritation in human operators in various domains such as online-based job information systems (Barber & Lucas, 1983), computer-based text file editing programs (Guynes, 1988), computer-based monitoring systems (Schleifer & Amick, 1989), basic human-computer interactions (Kuhmann, Boucsein, Schaefer, & Alexander, 1987; Szameitat, Rummel, Szameitat, & Sterr, 2009), and computer-based virtual reality environments (Allison, Harris, Jenkin, Jasiobedzka, & Zacher, 2001; Meehan, Razzaque, Whitton, & Brooks Jr, 2003). While time delay has been commonly addressed by improvements in technology, it is also regarded as an unavoidable, inherent component of teleoperation technology, telerobotics, and human-robot interaction (HRI) (Adelstein, Lee, & Ellis, 2003; Lum, Rosen, Lendvay, Sinanan, & Hannaford, 2009).

Studies that have explored the influences of time delay on teleoperation have focused on the negative consequences in task performance, such as decreased task accuracy (Owen-Hill et al., 2014) and increased task error rate (Szameitat, et al, 2009). However, the feelings of operators while experiencing time delay during teleoperation are rarely addressed compared to studies of overall human-robot joint task performance. Studies in general human-computer interaction have explored the emotional (affective) reaction to time delay, but the emotional states of operators are the least explored area in telerobotics. For example, lag in communication between humans and computers are known to cause increased heart rate and decreased perceptual stability (Allison, Harris, Jenkin, Jasiobedzka, & Zacher, 2001; Meehan, Razzaque,

Whitton, & Brooks Jr, 2003; Ash, Palmisano, Govan, & Kim, 2011). Studies have shown that the emotional state of an operator can affect productivity because human emotion is a significant factor in the quality of task performance (Picard, 2002; Klein, Moon, & Picard, 2002). If advances in teleoperation could be combined with human-robot interfaces that incorporate more consideration of empathy and affect, technology may move ever closer to authentically embodying the richness of the social interactions between humans.

Recent studies have explored affective communication between humans and robots, including the emotional expressions of robots (Kędzierski, Muszyński, Zoll, Oleksy, & Frontkiewicz, 2013), the effectiveness of the movements which robots use to convey emotional expressions (Nomura, & Nakao, 2010), and the development of artificial empathy of robots (Asada, 2014). These showed that emotion is becoming a meaningful factor in the design of HRI. Although studies have considered the emotional expressions of robots, the emotional experiences of human operators are difficult to establish due to the fluctuations in the emotional aspects of their physiological states (Yang & Dorneich, 2015).

Consequences of feedback from robots are important to examine because they trigger automatic emotional responses in human operators (Nass, Fogg, & Moon, 1996; Fogg, 2002). Consideration of the emotional, cognitive, physiological, and performance effects of time delay on human operators in teleoperation would enable the design of systems that could effectively assess and respond to operator emotions to ultimately yield improved human-robot joint performance. Moreover, detecting operator emotions is key to building affect-aware systems to mitigate the negative emotional states of operators and improve the overall productivity of teleoperation.

It is difficult to completely eliminate delayed responses between operators and robotic systems in modern technology (Prewett, Johnson, Saboe, Elliott, & Coovert, 2010; Owen-Hill, Suárez-Ruiz, Ferre, & Aracil, 2014). Establishing the link between time delay and its cognitive and physiological influences on human operators would be valuable for identifying the impact on human-robot joint performance. The next section discusses related work of time delay issues in teleoperation and characterizing human emotion. Based on an understanding of previous studies, an experiment was conducted to investigate the emotional, cognitive, physiological, performance, and usability effects of time delay while participants remotely navigated a robot vehicle in mazes of differing complexity. The experimental method focused on delayed response and its impact on operator emotions in order to support more empirical approaches in teleoperation. Results are described and discussed detailing the effect of time delay on operator emotions, cognitive workload, physiological arousal, performance, and usability in a human-robotic control task.

## Related Work

Several research areas are relevant to the study of time delay in human-machine interaction. First, the consequences of time delay in human-machine interaction are discussed to understand its attributes. Various aspects of human emotions are reviewed to understand their impact on human performance. Finally, since the evaluation will rely on measuring emotions, studies of the use of physiological measurements for emotional states are reviewed to assess existing methods for characterizing human emotions.

**Consequences of Time Delay in Human-Machine Interaction**

Various studies have looked at the consequences of time delay in human-machine interaction in relevant domains as teleoperation, human-computer interaction, and virtual environments. In teleoperation, work has focused on the performance and workload effects of time delay. Buffering time delay solely in the video stream of telemanipulation has led to decreased accuracy in a mock welding task with a robot arm (Owen-Hill et al, 2014). Moreover, time delay causes an increase in task completion times in telerobotic control of neutral buoyancy vehicles to simulate the microgravity environment of space (Corde Lane et al., 2002). A three-second delay caused a 132% increase in a free-flight maneuver task compared to the task completion time of a free-flight without time delay (Corde Lane et al., 2002). While analyzing human operators' aspects, time delay lowered the human operators' ability to recognize the environment through an unmanned ground vehicle, resulting in decreased task efficiency and observation sensitivity compared to when there was no time delay (Luck, McDermott, Allender, & Russell, 2006). In a similar manner, the effect of constant and random delays led to lower performance and higher operator workload in remote-controlled telerobot systems (Sheik-Nainar, Kaber, & Chow, 2005).

In the context of human-computer interaction, the effects of sporadic brief delays within a computer game (with an average duration of 1.6s) showed decreased performance, and increased reaction times and error rates (Szameitat et al., 2009). Short and long time delays in a simulated computer workplace system have been shown to have differential effects on operator performance and stress responses (Kuhmann, Boucsein, Schaefer, & Alexander, 1987). The participants under conditions of long (8 seconds) system response times, as compared to those under short (2 seconds) system response times, showed lower error rates in performance. In

physiological measurements, lower levels of systolic blood pressure were seen under long system response times, but a higher number of skin conductance reactions and pain symptoms (a headache, eye pain) were measured under short system response times (Kuhmann, Boucsein, Schaefer, & Alexander, 1987). In addition, lengthy system response times resulted in lower satisfaction and productivity among operators. Lower operator satisfaction may cause lower motivation to use current systems and drive operators to seek alternative devices. If there are no alternatives available to operators, they tend to adapt to the lags, but lower satisfaction can ultimately lead to lower productivity (Hoxmeier & DiCesare, 2000).

Due to the negative consequences of time delay, several studies intentionally used time delay to induce operator frustration. For example, delayed feedback was employed to introduce frustration while operators searched for given objects by using a keyboard and mouse (Klein, Moon, & Picard, 2002), and introducing time delay gaps between mouse input and screen output also effectively induced operator frustration (Powers, Rauh, Henning, Buck, & West, 2011).

Time delay was also investigated in virtual environment studies. When a virtual reality system had latency, the participants became significantly more likely to experience oscillopsia. This shows that perceptual instability arises with increased display lag in virtual environments (Allison, Harris, Jenkin, Jasiobedzka, & Zacher, 2001). In addition, display delay during active head oscillation impairs the illusion of vection in a virtual environment, which illustrates that viewers feel like they have moved and yet the world is stationary (Ash, Palmisano, Govan, & Kim, 2011). System latency has led to decreased perceptual sensitivity in virtual environments. The required length of latencies is less than 16 ms in order to maintain the same level of perceptual stability as a virtual environment without latency (Ellis, Mania, Adelstein, & Hill, 2004).

**Human Emotion in Human-Machine Interaction**

The quality of interaction between a human and a machine can be influenced by human emotions. Generally, humans expect appropriate reactions from machines because human operators accept machines as a team member like other human co-workers (Nass, Fogg, & Moon, 1996). Ferdig and Mishra (2004) demonstrated that humans felt just as happy when their partner treated them fairly and they felt just as unhappy and angry when their partner betrayed them, regardless of if that partner was a human or a computer. Furthermore, the participants continually tried to read intentionality into the reactions of their partners.

Operators preferred a system where in addition to text-based interaction with a computer, they were able to report their feelings to it (Klein et al., 2002). Several experiments have shown that the effect of being persuaded by a computer had the same effect as being convinced by another human (Nass, Moon, Fogg, Reeves, & Dryer, 1995; Nass et al., 1996; Fogg, 1998; Fogg, 2002). Interaction with computing devices may elicit an emotional response from the user if the device causes people to perceive it as a social actor. Humans naturally respond to social presences emotionally (e.g., feeling of empathy or anger) with social norms (e.g., turn-taking rules) (Nass et al., 1995).

Human emotions, especially negative emotions (e.g., frustration), have been considered a significant factor influencing task performance in various fields. For instance, early psychological research demonstrated that frustration reduced the performance quality of an intelligence test (Waterhouse & Child, 1953), reduced perceptual-motor performance in children (Solkoff, Todd, & Screven, 1964), and reduced employee performance (Spector, 1975). Frustration, defined as an emotional state in which obstacles block the possibility of achieving a goal (Lawson, 1965), is a complex emotion related to anger and disappointment; aggression is

one of its consequences (Dollard, Miller, Doob, Mowrer, & Sears, 1939). In human-computer

interaction, frustration is one of the most common experiences for computer operators (Ceaparu,

Lazar, Bessiere, Robinson, & Shneiderman, 2004). Human society is becoming more reliant on

computers and robots to accomplish tasks; nearly all aspects of our work and private lives need

assistance from technology in one form or another. Despite the ongoing technological

innovations, frustration remains a universal problem for humans operating robots, personal

handheld devices, automated systems, or any other computer system. As such, frustration has

become a significant interest in the context of human-machine interaction. Frustration stemming

from the use of computers causes operators to waste an average of 42-43% of their time (Lazar et

al., 2005). In a recent study, Hertzum (2010) reported that operators spent an average of 16% of

their time on fixing the problems they encountered and 11% on redoing lost work.

Frustration is a topic worth exploring for reasons other than its relation to productivity.

Sensing and responding to emotions such as frustration would enable the development of a more

human-like affective computer (Picard, 2002). In fact, technological advances have sparked the

emergence of conversation interfaces (Oviatt, 2004). In this context, understanding operator

frustration and determining its causes are of crucial importance.

**Characterizing Human Emotion**

Human emotions can be characterized by arousal and valence. Arousal refers to

emotional excitedness or activation, and ranges from calming or soothing to exciting or agitating.

Valence refers to whether the emotional state of an operator is positive or negative, and ranges

from highly positive to highly negative (Schlosberg, 1954; Russell, 1980; Frijda, 1986; Lang,

Greenwald, Bradley, & Hamm, 1993; Kensinger, 2004). These two aspects of human emotion

can be measured (Lang, Greenwald, Bradley, & Hamm, 1993; Kim, Bang, & Kim, 2004; Nasoz, Alvarez, Lisetti, & Finkelstein, 2004; Li, & Chen, 2006).

In HRI, human emotion can be measured as an input. Montagne et al. (2007) developed the emotion recognition task, which is a computer-generated paradigm for measuring six basic facial emotional expressions: anger, disgust, fear, happiness, sadness, and surprise. Happiness was the easiest emotion to recognize by the computer. Sadness and fear were the most difficult emotions to recognize because sadness involves subtle changes in the face, and fearful faces are rarely encountered in everyday life.

Various sensors have been used to assess emotional states, including palmar sweat, event-related brain potentials, electroencephalography, heart rate, pupil diameter, muscle tension, electromyography (EMG), cortisol levels, respiration, blood volume pulse (BVP), video (facial expressions and gestures), and galvanic skin response (GSR) which includes electrodermal activity (EDA) (Kiesler, Zubrow, Moses, & Geller, 1985; Kramer, 1991; Wiethoff, Arnold, & Houwing, 1991; Scheirer, Fernandez, Klein, & Picard, 2002; Octavia, Raymaekers, & Coninx, 2011).  One of the most commonly used measurements to detect human emotion is facial expression. Facial expressions are the results of muscle movements; thus patterns of muscle activation can represent emotional states (Ekman, 1970). Recording and analyzing an individual's facial expressions have been used to assess anger, happiness, sadness, surprise, dislike, and fear (De Silva, Miyasato, & Nakatsu, 1997; De Silva & Ng, 2000), and disgust (Benţa et al., 2009). For instance, FaceReader is a video-based system that automatically analyzes facial expressions to calculate arousal and valence. It detects 'happy' as mid-arousal and positive valence, and 'angry' as high arousal and negative valence (Loijens et al., 2012). Another assessment of facial expressions is the use of a facial electromyography (fEMG) sensor.

The fEMG is used for measuring a human's emotional state by placing a sensor at the corrugator of the face (Hazlett, 2003). The fEMG sensor measures the electrical activity produced by facial muscles. The sensitivity and accuracy of the fEMG are dependent on where the electrode is placed over the muscle of interest.

Skin conductance has been frequently employed to detect human emotion. It is the basis of many sensor techniques, including electrodermal response, psychogalvanic reflex, skin conductance response, and skin conductance level (Conesa, 1995). Studies have found that skin response is positively correlated with body arousal (Meisner, Isler, & Trinkle, 2008; Swangnetr, Zhu, Kaber, & Taylor, 2010; Zoghbi, Croft, Kulic, & Van der Loos, 2009; Rosenthal-von, Krämer, Hoffmann, Sobieraj, & Eimler, 2013; Chen, King, Thomaz, & Kemp, 2014). For instance, Bradley and Lang (2000) found a significant correlation between emotional stimuli and skin conductance. Sweat gland activity raises skin conductance as a result of the sympathetic branch of the autonomic nervous system becoming highly aroused (Carlson, 2013). This activated skin conductance is associated with body arousal and related to the emotional response of an individual (Lang, 1995; Boucsein, 2011).

Electrodermal activity (EDA) is defined as changes in the skin's electricity and considered to be a similar signal to galvanic skin response. An EDA sensor can detect autonomic changes in the electrical properties of the skin. Fluctuations and reductions in EDA signals were observed when subjects were reminded of negative memories (Barrowcliff, Gray, Freeman, and MacCulloch, 2004). In contrast, higher EDA signals were detected when participants were shown evocative photos (Radin, 2004). Moreover, EDA sensors were employed to investigate the implicit emotional responses that arise without conscious awareness or cognitive intention, such as threat, salience, and novelty (Braithwaite, Watson, Jones, & Rowe, 2013).

In terms of HRI, skin conductance was used to estimate human affective states. Skin conductance was used as an indicator of affective arousal of the participants while they observed the robot's motions, which were called "robot body language." (Kulic & Croft, 2007). Skin conductance, heart rate, and facial muscle contractions were used to interpret how operators reacted to a robot's body language (Kulic & Croft, 2007).

## Method

### Objective

The objective of this study was to examine how delayed responses from the robotic system impact operator emotional response, cognitive workload, task performance, physiological arousal, and usability. The results were used to establish an integrated understanding of the influences of time delay on teleoperation.

### Hypothesis

Time delays in human-robot interaction will cause increases in a human operator's cognitive workload, frustration, aggression, and physiological arousal while decreasing task performance, and system usability. These effects are expected at both low and high task difficulty levels.

### Participants

A priori power analysis indicated that a sample size of 20.93 would be sufficient to detect a significant effect of independent variables with a power of .90 and an alpha of .05 based on statistical power for a two-way, within-subjects ANOVA. A total of 21 university students (14 males, 7 females) participated in the experiment. The participants' average age was 28.3 (range: 22 – 43). All subjects were experienced computer operators who have been currently using

computers an average of 3.8 hours in a day. All participants had a normal or corrected-to-normal vision, excluding the possibility of diminished attention due to vision problems.

**Task**

Participants were asked to conduct two tasks: target search and alert detection. Participants were asked to conduct both tasks simultaneously during the experiment.

*Target Search.* The task consisted of controlling a robot vehicle (Zhong, 2013) from a remote location via a joystick and navigating it through a maze. The robot vehicle was equipped with a video camera. Participants could only see the video feed displayed on a monitor as he or she navigated the robot. While navigating two different mazes, participants were asked to identify as many identical cylindrical objects placed throughout the maze as they could. Participants were asked to verbally report the objects as either "new" or "old," depending on if they had encountered this object previously. Since all objects have same size, shape, and color, participants had to remember the structure and path inside of maze in order to identify whether targets were new or old.

*Alert Detection.* This task required the participants to pull the trigger on the joystick when they heard audio beeps while completing the other tasks. The beeps occurred at random intervals that averaged 30 seconds between beeps.

**Independent Variables**

The two independent variables were *Time delay* (no time delay, time delay) and *Task difficulty* (low, high). Time delay was elicited via the introduction of feedback delay in control inputs. Task difficulty was manipulated via the complexity of the maze.

*Time delay.* In the time delay condition, participants experienced a lag between the time they input a control command to the robot, and the time the robot responded. The lags occurred

randomly on some but not all of the control inputs, averaging approximately 10 delays per minute. In the no-time delay condition, the system responded to operator commands with no perceivable delays.

The length and timing of the delays in the time delay condition were determined via a combination of drawing on previous literature and on pilot studies. The goal was to determine what length of delay would elicit task difficulty without causing the operator to simply give up on the task. Previous literature has defined four categories of time delay ranges based on human perception (Miller, 1968; Card, Robertson, & Mackinlay, 1991). Operators sense that the system is responding instantaneously for time delays of 0.1 seconds or less. Time delays between 0.1 - 1.0 second are recognizable to the operator, but do not cause a loss of the feeling of operating smoothly. Generally, no special make-up feedback is necessary. Time delays between $1.0 - 10$ seconds cause noticeable delay, but the operator is still able to keep their attention focused on the task. For delays longer than 10 seconds, operators can become distracted and shift attention to other tasks while waiting for the system to finish (Miller, 1968; Card, Robertson, & Mackinlay, 1991). Based on this, a pilot study was conducted to test the effects of time delay between 0.0 - 5.0 seconds. Five participants (average age 30.4 years, ranging from 24 to 45) remotely controlled the robot vehicle in the same experimental setting as the main study and self-reported the levels of task load they experienced using NASA Task Load Index (TLX) (Hart & Staveland, 1988). The results of the NASA TLX were the highest for a delay duration between $2.0 - 3.0$ seconds. A longer duration of delay (4.0s and 5.0s) led some of the participants to assume the system was not in working order and caused them to discontinue the task. Based on these findings, the duration of each feedback delay was randomly set at either 2.0 or 3.0 seconds.

*Task difficulty.* Two different complexity levels of mazes (see Figure 2) were used in order to manipulate task difficulty of the navigation task. A simple maze was used for low task difficulty, and a complicated maze was used for high task difficulty. The same number of red-colored targets (14) were present in both mazes. To reduce any learning effect, the following steps were taken: participants started each maze from different entrances, and between each trial, the locations of targets were changed and the mazes were rotated.
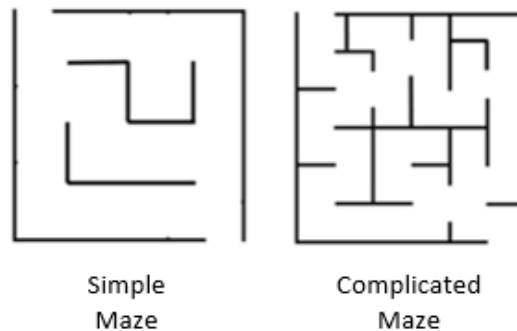


**Figure 2. Floor plans and top view of the simple and complicated maze.**

## Dependent Variables

The dependent variables were physiological arousal, emotional state, cognitive workload, task performance, and usability. Except physiological arousal and usability, each had one subjective and one objective method of measurement. In addition, participants were asked to rate their perceived task difficulty in order to verify that the manipulation of the maze difficulty produced the desired effect of making the tasks more difficult for the complicated maze than the simple maze. The dependent variables are described in Table 1.

**Table 1. The measurements for both independent variables verification and dependent variables.**

| Variables | Metric | Measurement (Unit) | Frequency | Data Type | |
|---|---|---|---|---|---|
| Independent Variable Manipulation Verification | Perceived Task Difficulty | Likert Scale 1 – 5 | After each trial | Subjective | Ordinal |
| Emotional State | Emotional Rating (7 emotions) | Likert Scale 1 – 5 | After each trial | Subjective | Ordinal |
| | FaceReader Emotion Distribution (7 emotions) | 0-100% | During each trial | Objective | Continuous |
| Cognitive Workload | TLX Subscale Mental Demand | Scale 0 – 10 | After each trial | Subjective | Ordinal |
| | Reaction Time | Second (s) | Nine times per trial | Objective | Continuous |
| Task Performance | TLX Subscale Performance | Scale 0 – 10 | After each trial | Subjective | Ordinal |
| | Targets correctly identified | 0 – 100 (%) | After each trial | Objective | Ordinal |
| | Total unique targets found | 0 – 100 (%) | After each trial | Objective | Ordinal |
| Usability | Perceived Task Difficulty | Likert Scale 1 – 5 | After each trial | Subjective | Ordinal |
| | Perceived Speed | Likert Scale 1 – 5 | After each trial | Subjective | Ordinal |
| | Perceived Smoothness | Likert Scale 1 – 5 | After each trial | Subjective | Ordinal |
| | Appropriateness | Likert Scale 1 – 5 | After each trial | Subjective | Ordinal |
| | Operator Satisfaction | Likert Scale 1 – 5 | After each trial | Subjective | Ordinal |
| Physiological Arousal | Electodermal Activity (EDA) | Microsiemens (μS) | During each trial | Objective | Continuous |

*Independent Variable Manipulation Verification* – Perceived Task Difficulty. Participants were asked after every trial their subjective rating of task difficulty (on a 5-point Likert scale). Task difficulty was an independent variable manipulated by varying the complexity of the maze; thus, it was expected that perceived task difficulty would be higher for the complex maze than the simple maze. This would verify that the independent variable manipulation was successful.

*Emotional State.* The relationship between time delay and human emotions was measured in two ways: 1) subjectively through a questionnaire and the NASA TLX, and 2) objectively with FaceReader, an affect sensor that classifies seven different emotions based on video recognition of facial expressions.

The questionnaire contained a five-point Likert-type scale from 'never' to 'all of the time' about seven different emotions (happy, angry, sad, surprised, scared, disgusted, and frustrated). The frustration subscale of NASA TLX was also employed to gather data of self-rated frustration level. Both were administered after every trial.

FaceReader models the face with over 500 points and recognizes patterns against by a large database of annotated images (Cootes and Taylor, 2000). FaceReader was chosen for sensing emotions for four reasons. Firstly, facial-expression-based FaceReader has been shown to be an accurate assessment of emotion (Den Uyl & Van Kuilenburg, 2008; Loijens et al., 2012). Secondly, FaceReader detects the facial expressions in real-time (an important requirement for planned follow-on work). Another facial recognition software (Affdex) was considered, but it does not provide a real-time detection. Thirdly, FaceReader is unobtrusive as it only uses a webcam rather than a sensor attached to the face. Lastly, FaceReader has an ability to detect and classify wide-range of emotions: Paul Ekman's six universal emotions and neutral state (Ekman, 1970), whilst Affdex has only four emotional categories. FaceReader outputs a value between 0 and 100% for each of the seven emotions it can classify (Loijens et al., 2012), where the total across all seven emotions equals 100%.

*Cognitive Workload.* Reaction time in the alert detection task was used to objectively assess workload (Patten, Kircher, Östlund, & Nilsson, 2004). The post-trial mental demand subscale scores reported on the NASA TLX served as the subjective measurements of workload.

*Task Performance.* The scores of task performance were computed by using two methods: the percentage of targets correctly identification percentage and percentage of the total unique targets found. The correct identification percentage denoted the number of correct classification ("new" or "old") of targets seen by the participants. The unique targets found percentage is the number of unique targets found of the total number of unique targets in the maze. The NASA TLX performance subscale scores served as a subjective measure of performance.

*Usability of the Robot Control System.* After each trial, participants were asked to complete a post-trial survey which included five questions about controlling the robot in the areas of difficulty, speed, smoothness, appropriateness of feedback, and satisfaction. A post-experiment asked participants to describe the most difficult and the easiest parts of the robotic control task, the strategies they used to control the robot, and the three most frustrating aspects of the robotic control task.

*Physiological Arousal.* An electrodermal activity (EDA) sensor has been shown to be a useful indicator of emotional arousal since it can measure the activation of the sympathetic nervous system independent of the cause. The target identification task required high attention and memory load, which should lead to a change in the rate of body-circulating adrenaline to activate the sympathetic nervous system in human operators (Squire, 1987; Gross, 1998). Although facial EMG sensors might provide an alternative way to detect human emotion, facial EMG sensors only measure muscle movements of the surface of the human face which indicate the level of valence (Mata-Cervantes, Westerman, Burke, Hill, Wyatt, 2014), and do not directly detect the level of arousal which includes changes to the sympathetic nervous system such as sweat, body temperature, and skin conductance. In addition, the EDA sensor is deployed in a watch-shaped wrist sensor and thus does not require wired facial sensors that would interfere with the facial expression of participants. The combination of a facial recognition-based FaceReader and wrist-based EDA sensor allowed collection of both emotion (FaceReader) and arousal (EDA) simultaneously. Signal data from EDA sensor were gathered during the trials to measure body arousal. EDA data were calculated by comparing the signals from each trial with baseline data on a per participant basis.

**Experimental Design**

This experiment was a 2 (*time delay*: no time delay vs. time delay) x 2 (*task difficulty*:

low vs. high) within-subjects, repeated measures design. Each condition was tested twice per

participant (i.e., two replications); thus, each participant completed a total of eight trials. In each

trial, the participant conducted both the target search and alert detection tasks concurrently. The

order of each combination of the two independent variables was counterbalanced across

participants using a series of 8x8 Latin squares (for the 8 trials) to account for any learning

effects (Table 2). Each trial had a five-minute time limit.

**Table 2. Counterbalanced trials. Each participant encountered the independent variable conditions in a different order.**

| Trial | Participant 1 | Participant 2 | Participant 3 | Participant 4 | ... |
|---|---|---|---|---|---|
| Trial 1 | No Delay Complicated Maze | Delay Simple Maze | Delay Complicated Maze | No Delay Simple Maze | ... |
| Trial 2 | Delay Simple Maze | No Delay Complicated Maze | No Delay Complicated Maze | Delay Complicated Maze | ... |
| Trial 3 | Delay Complicated Maze | No Delay Simple Maze | No Delay Simple Maze | No Delay Complicated Maze | ... |
| Trial 4 | No Delay Simple Maze | Delay Complicated Maze | Delay Simple Maze | Delay Simple Maze | ... |
| Trial 5 | Delay Simple Maze | No Delay Complicated Maze | No Delay Simple Maze | Delay Complicated Maze | ... |
| Trial 6 | No Delay Complicated Maze | Delay Complicated Maze | No Delay Complicated Maze | Delay Simple Maze | ... |
| Trial 7 | No Delay Simple Maze | Delay Simple Maze | Delay Complicated Maze | No Delay Simple Maze | ... |
| Trial 8 | Delay Complicated Maze | No Delay Simple Maze | Delay Simple Maze | No Delay Complicated Maze | ... |

**Procedure**

The experiment began with the consent process, a short briefing, a pre-survey, and the

attachment of physiological sensors on the participant. All study materials can be found in

Appendix A. The participants were asked to read a magazine or a book for 15 minutes in order to

collect baseline EDA data for each participant. Participants were trained to remotely operate the

robot, including the correct way to hold the joystick and how to navigate the robot. Then they

had approximately 30 minutes to practice navigating the robot. When the participants were able to smoothly navigate the robot, they began the trials.

Participants completed eight trials. Between trials, participants were asked to complete a post-trial survey and the NASA TLX. Video of participants' facial movements was collected in order to analyze their emotional statuses. After finishing all eight trials, the participants were asked to fill out a post-experiment survey to gather their opinions, strategies, and ideas to improve future studies. During the debriefing, the experimenter explained to the participants that the true goal of the study was to study the effect of time delay, as they had been initially told that the study was intended to test the robot control software.

**Testing Apparatus**

Participants controlled the robot via a joystick (see Figure 3a). They were spatially separated from the maze and could only see the video provided from the robot's camera feed to the monitor screen (Zhong, 2013). A physiological sensor was employed for this experiment, which was EDA. It was placed around the wrist of the non-dominant hand (see Figure 3b). The EDA sensor used was an Affectiva Q-sensor connected to a Dell Precision T1700 desktop PC.



(a)                (b)                (c)                (d)
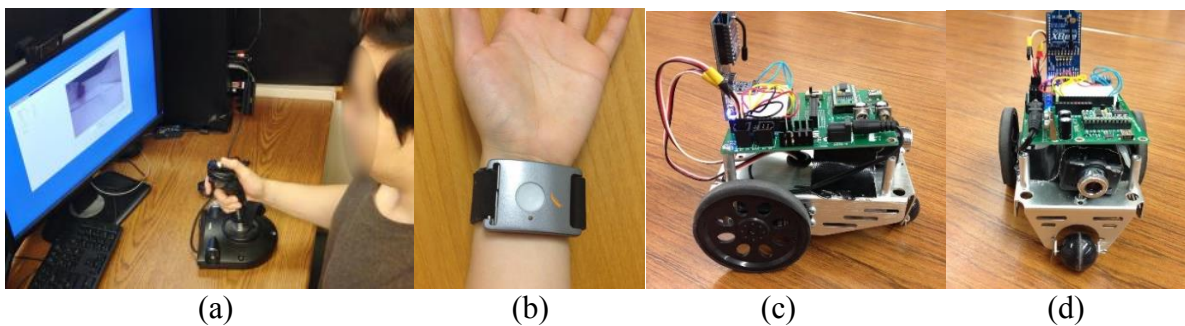
**Figure 3. (a) Experimental setup with a joystick. (b) EDA sensor around the wrist of the non-dominant hand side during testing. (c) Robot's side. (d) Robot's front.**

**Data Analysis**

The Shapiro-Wilk test was employed to check normality of data and Bartlett's test was used to test the homogeneity of variance. A multivariate, two-way, within-subjects ANOVA was used to analyze the dependent variables. Measured data included ratings of emotional states and usability, TLX survey, target identification correct score, and reaction time. Results are reported as highly significant for a significance level alpha <.001, significant for alpha <.05, and marginally significant for alpha <.10. Additionally, each participant's EDA signal data was baselined and then all participant's data was averaged to create profiles of signals in different conditions.

**Limitations and Assumptions**

EDA directly measures physiological states associated with arousal rather than the emotional states that cause arousal, implying that EDA may only indirectly measure the exact emotional activation of the participants.  Since the task was developed to navigate mazes by using only a restricted view from a robot, it required an ability to recognize paths through the monitor. The degree of such perceptual ability, however, varies from individual to individual and might have influenced the task outcome.

<div align="center">Results</div>

**Independent Variable Manipulation Verification – Task difficulty**

Participant task difficulty ratings were compared in the no-time delay condition in order to verify that the simple and complicated maze induced low and high task difficulty, respectively. Participants felt that the complicated maze (M=3.00, SD=0.91) was significantly (F(1,20)=22.02, *p<.0001*) more difficult than the simple maze (M=2.61, SD=1.02). These results

confirm that the varying levels of maze complexity successfully caused significantly different levels of task difficulty. In addition, time delay also significantly (F(1,19)=37.56, *p<.0001*) increased task difficulty (Figure 4). However, the interaction between the time delay condition and the structure of the maze was not significant.
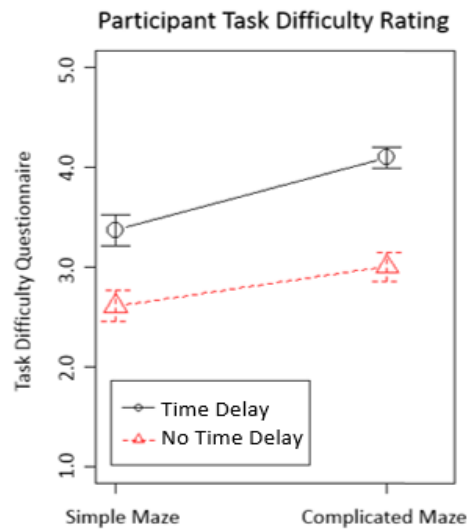


**Figure 4. The mean and standard error of task difficulty.**

**Emotional State**

*Subjective Emotional Rating.* Of the seven emotions that participants rated via post-trial questionnaires, only frustration and anger showed significant results. Participants' subjective rating of frustration showed that the feedback delay in both the simple (M=2.98, SD=1.25) and complicated mazes (M=3.19, SD=1.29) was significantly (F(1,19)=30.37, *p<.0001*) higher than no feedback delay of both the simple (M=1.95, SD=0.89) and complicated mazes (M=2.50, SD=1.06) (see Figure 5a). The effect of task difficulty on operator frustration was also significant in time delay condition (F(1,20)=6.33, *p<.03*).

In addition, the NASA TLX frustration subscale provided similar results in that the feedback delay of both simple (M=6.06, SD=2.11) and complicated mazes (M=6.61, SD=2.01)

leads to significantly (F(1,20)=89.11, *p<.0001*) higher frustration than no feedback delay of both

the simple (M=3.11, SD=2.22) and complicated mazes (M=3.94, SD=2.39) (see Figure 5b). The

effect of task difficulty on the NASA TLX frustration subscale was also significant in time delay

condition (F(1,20)=6.11, *p<.03*).



(a)                                        (b)
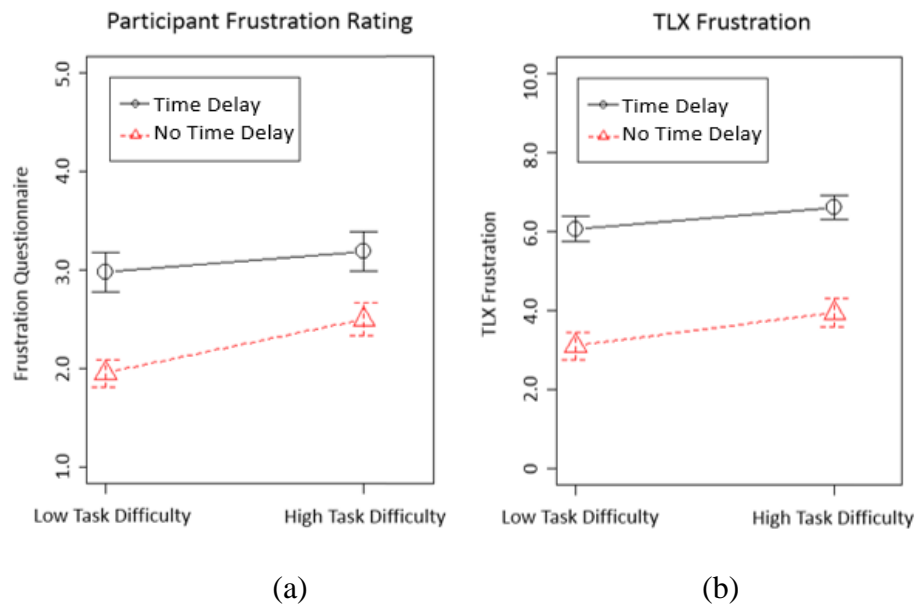
**Figure 5. (a) The mean and standard error of the (a) frustration questionnaire results and (b) the TLX frustration results.**

The time delay condition led to significantly higher anger ratings (F(1,19)=29.179,

*p<.0001*) with both low (M=2.15, SD=1.22) and high task difficulty (M=2.33, SD=1.26) than the

no time-delay condition with both low (M=1.27, SD=0.55) and high task difficulty (M=1.69,

SD=1.05) (see Figure 6a). However, task difficulty and the interaction effect were not

significant.

*FaceReader Emotion Distribution.* Of the seven emotional classifiers, only anger showed

significant results (F(1,20)=5.13, *p<.05*). The average intensity value of anger in both low

(M=0.53, SD=0.24) and high task difficulty (M=0.57, SD=0.29) was higher than no time delay

with both low (M=0.41, SD=0.25) and high task difficulty (M=0.46, SD=0.24) (see Figure 6b).

However, task difficulty and the interaction between task difficulty and time delay were not significant.



(a)                                              (b)
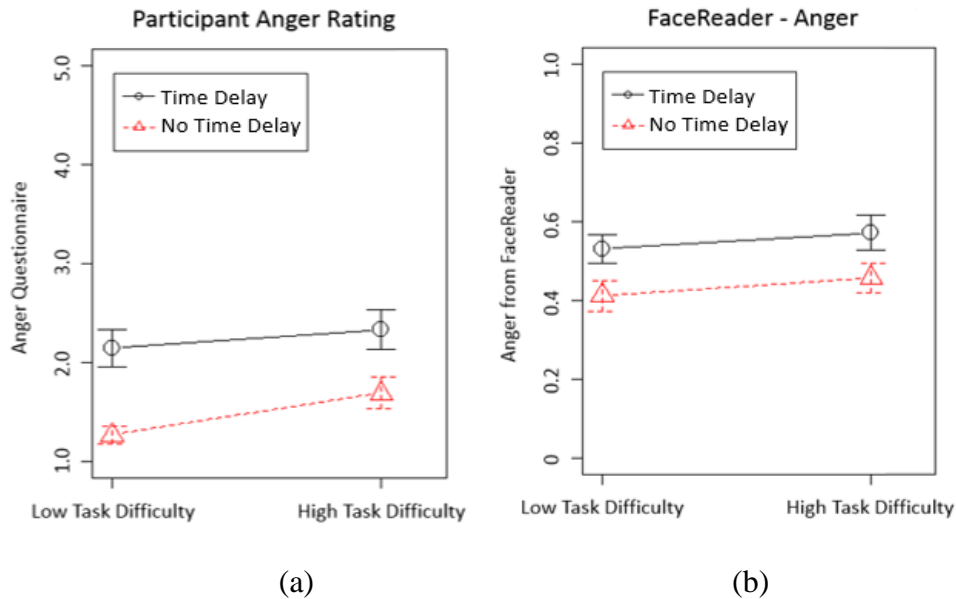
**Figure 6. (a) The mean plot and standard error of anger questionnaire results. (b) The mean plot and standard error of anger from FaceReader analysis.**

**Cognitive Workload**

*TLX Mental Demand.* Time delay condition of both low (M=6.21, SD=1.85) and high task difficulty (M=7.73, SD=1.28) significantly (F(1,20)=49.9, *p<.0001*) led to higher mental demand when compared to the no time-delay condition of both low (M=4.43, SD=1.97) and high task difficulty (M=5.44, SD=1.49) (see Figure 7a). The task difficulty also significantly (F(1,20)=17.94, *p<.0005*) increased mental demand. The interaction effect of time delay and task difficulty was not significant.

*Reaction Time.* Time delay of both low task difficulty (M=0.96, SD=0.67) and high task difficulty (M=1.13, SD=0.88) significantly (F(1,18)=9.03, *p<.008*) led to slower reaction time than the no time delay of both low (M=0.82, SD=0.38) and high task difficulty (M=0.93, SD=0.56). Participants reacted more slowly from external stimulus with the feedback delay of

the system (see Figure 7b). The task difficulty also significantly ($F(1,18)=6.52$, *p<.03*) decreased

reaction time in the complicated maze compared to the simple maze. However, the interaction

between frustration and task difficulty was not significant.



(a)                                                              (b)
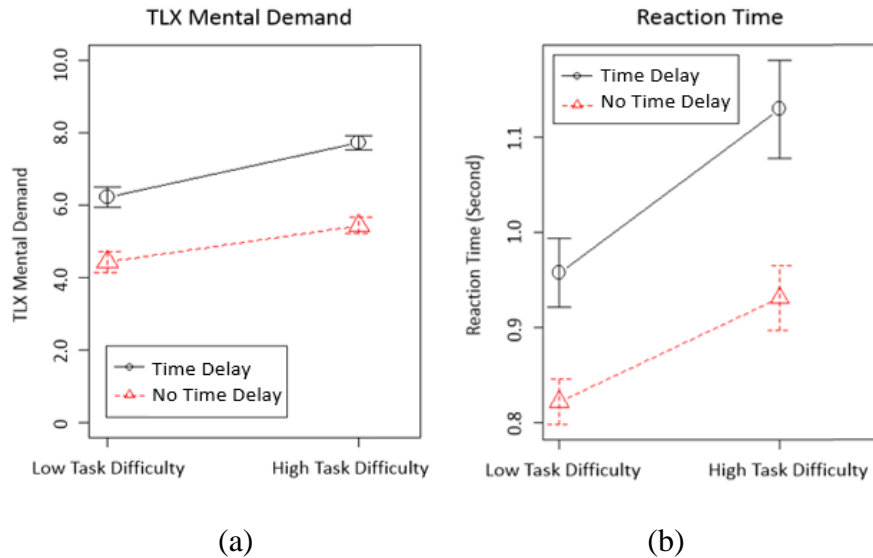
**Figure 7. The mean and standard error of (a) reaction time and (b) TLX mental demand.**

## Task Performance

*Targets correctly identified.* The effect of time delay was marginally significant

($F(1,20)=3.87$, *p=.0632*) on the correct percentage of target identification. Participants tended to

more correctly identify the objects when they navigated in the no time-delay condition with both

low (M=86.7, SD=13.4) and high task difficulty (M=80.4, SD=2.92) than in the time delay

condition with low (M=80.4, SD=18.7) and high task difficulty (M=77.1, SD=4.10) (see Figure

8a). However, the effects of task difficulty and the interaction between time delay and task

difficulty were not significant.

*Total unique targets found.* Participants found significantly ($F(1,20)=44.06$, *p<.0001*)

more unique targets when they navigated in the no time-delay condition with both low (M=76.5,

SD=13.7) and high task difficulty (M=52.2, SD=18.1) than in the time delay condition with low

(M=60.5, SD=13.1) and high task difficulty (M=33.6, SD=13.5) (see Figure 8b). In addition, the

effect of task difficulty was significant (F(1,20)=77.13, $p<.0001$), where higher task difficulty

led to a lower percentage of unique targets found. However, the interaction between time delay

and task difficulty was not significant.

*TLX Performance.* The effect of time delay was significant (F(1,20)=37.27, $p<.0001$);

time delay condition with both low (M=5.82, SD=2.21) and high task difficulty (M=6.21,

SD=2.19) led to lower performance than no time-delay condition with both low (M=3.45,

SD=2.18) and high task difficulty (M=4.64, SD=2.13). The effect of task difficulty was

significant (F(1,20)=4.88, $p<.04$), where higher task difficulty led to a higher score of TLX

performance (see Figure 8c). However, the interaction between time delay and task difficulty

was not significant.



(a)                              (b)                              (c)
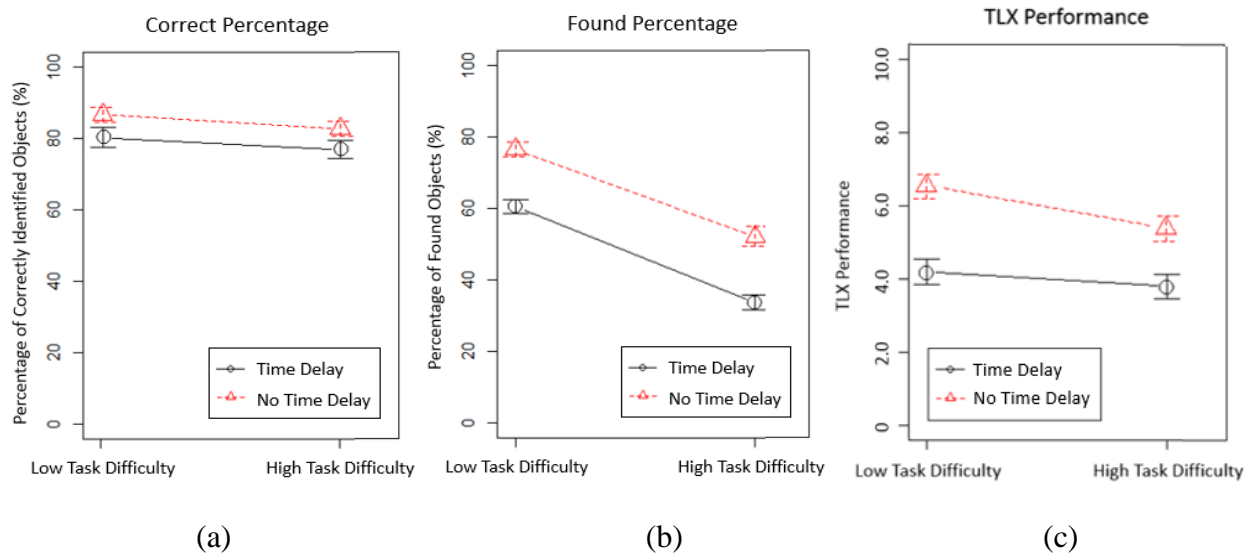
**Figure 8. (a) The mean plot and standard error of correct percentage of identified objects. (b) The mean plot and standard error of percentage of found objects. (c) The mean plot and standard error of TLX performance results.**

**Usability**

Table 3 shows that the results of four different questions about each trial: speed of control, smoothness of control, appropriateness of control, and operator satisfaction. The effect of time delay was significant for speed of control ($F(3,162)=15.20$, $p<.0002$), smoothness of control ($F(3,162)=33.00$, $p<.0001$), appropriateness of control ($F(3,161)=9.20$, $p<.003$), and operator satisfaction ($F(3,162)=23.47$, $p<.0001$). Task difficulty was marginally significant for the speed of control ($F(3,162)=10.59$, $p<.075$) and significant for appropriateness of control ($F(3,161)=5.50$, $p<.003$). However, the interaction between time delay and task difficulty was not significant for any of the measures.

**Table 3. Means and standard deviations of usability questionnaire. Conditions marked with a "*" were significant, "**" were highly significant, and "m" were marginally significant.**

| Usability | Low Task Difficulty: Mean (Std Dev) | | High Task Difficulty: Mean (Std Dev) | | *p*-value | |
|---|---|---|---|---|---|---|
| | No Time Delay | Time Delay | No Time Delay | Time Delay | Time Delay | Task Difficulty |
| Speed | 3.24 (0.83) | 1.90 (0.70) | 3.05 (0.99) | 1.69 (0.68) | ** | m |
| Smoothness | 3.14 (0.89) | 1.95 (0.67) | 3.14 (0.93) | 1.59 (0.54) | ** | Not Significant |
| Appropriateness | 3.61 (0.77) | 2.37 (0.83) | 3.34 (0.96) | 1.83 (0.76) | ** | ** |
| Satisfaction | 3.68 (0.72) | 2.39 (0.97) | 3.24 (0.88) | 2.09 (0.82) | ** | Not Significant |

**Physiological Arousal**

*Electrodermal Activity.* Only 16 of the 21 participants' sensor data were used due to a loss of five participants' data (caused by a poor fixation of the sensor to the wrist which resulted in a loose connection when the participants were moving). The EDA data was calculated by subtracting the participants' baseline data from the data collected during the trial. The effect of time delay was significant ($F(1,16)=2.62$, $p<.05$); time delay condition with both low ($M=1.92$, $SD=0.23$) and high task difficulty ($M=2.74$, $SD=0.15$) led to higher EDA than no time-delay condition with both low ($M=1.63$, $SD=0.19$) and high task difficulty ($M=2.51$, $SD=0.24$) (Figure 9). Moreover, the effect of task difficulty was significant ($F(1,16)=4.59$, $p<.0001$), where EDA increased for the more complicated task when compared to the simple task. However, the

interaction between time delay and task difficulty was not significant. In addition, Figure 10 illustrates the (averaged) EDA for the four conditions over the course of five minutes. The collected EDA data showed a marked increase in the time delay over the no time-delay condition, across both task difficulty conditions.
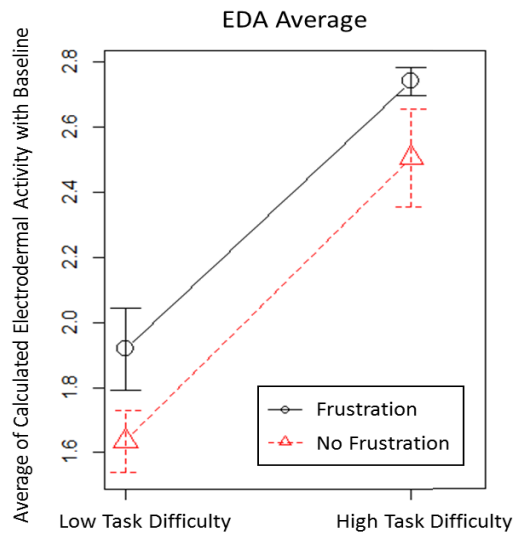


**Figure 9. The mean and standard error of EDA.**
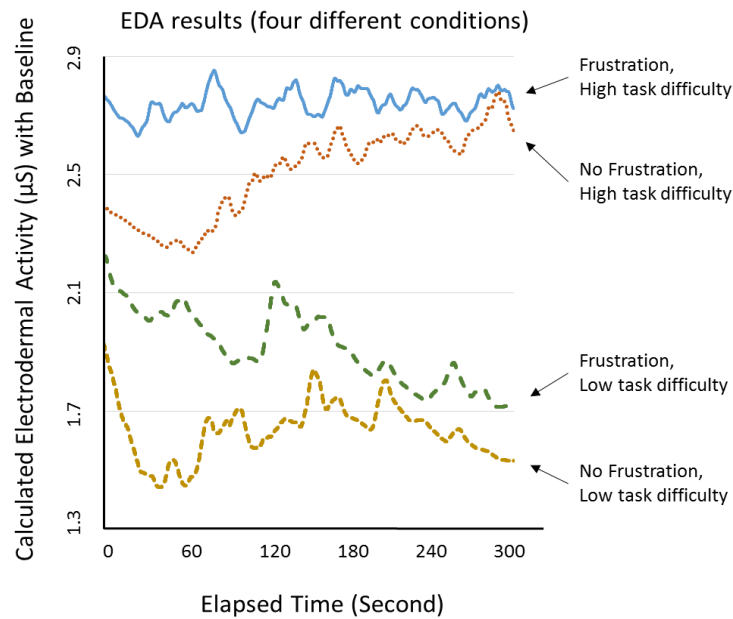


**Figure 10. Average change of electrodermal activity signals.**

Discussion

The results of the study show that time delay significantly influences physiological arousal, emotional states (frustration, anger), cognitive workload, and task performance. These results were true for both low and high task difficulties. Participants' reaction times to interruptions were slower, and they subjectively rated their mental demand higher in the time delay condition. In the post-experiment questionnaires, participants reported that they felt higher frustration and anger with the time delay than the no-delay condition. Participants experienced the lowest workload when they navigated without any delayed feedback.

Averaged results of EDA from 16 participants showed different levels of arousal subject to the combinations of time delay and task difficulty. The EDA profile showed that task difficulty represented a major factor in the level of arousal. The average change of EDA signals showed that task difficulty had a larger influence on the level of arousal when compared to time delay. The EDA results also showed that the presence of time delay raised the average level of arousal when compared to no time delay; this effect was present for both task difficulty levels. This was demonstrated in both the participants' subjective rating of task difficulty across the easy and complex maze, as well as demonstrated by the EDA results. Cognitive workload, reaction time, and performance also followed the same pattern, where both task difficulty and time delay were significant factors. This is in line with expectations from previous work. For emotional states however, time delay was significant, but task difficulty was not. While higher task difficulty has a significant impact on performance, workload, and arousal, it does not strongly affect emotional state. The presence of time delay, however, leads to an emotional response in addition to all the other impacts previously discussed. This has implications for the

design of systems that should minimize the frustration factor in order to properly calibrate user expectations for a given task difficulty.

In spite of steady time delay events throughout the trial, the EDA signals gradually decreased in low task difficulty conditions after the middle of the trial. It is possible that the participants acclimatized to time delay during the trials. If this is the case, then this has implications for the ability of EDA to detect arousal during long-term exposure. This is an area of further work that might establish a relationship between skin conductance and emotional excitedness or activation over long periods of time.

Among the seven emotions, only frustration and anger show a significant relationship with time delay. Frustration has been related to aggression (Dollard, Miller, Doob, Mowrer, & Sears, 1939; Miller, 1941; Morlan, 1949). In this study, we found that time delay significantly affected frustration and anger when measured objectively (FaceReader) and subjectively (questionnaire). During the trials with time delay, comments from participants included: "It is annoying," "I don't understand why it is not moving properly," "I'm almost angry." Anger is classified as having higher emotional activation than frustration (Barrett, 2006; Lindquist, & Barrett, 2008). If operators experience anger during their interactions with computers, their mental workload would be higher than in a neutral mood. As a consequence, human performance, productivity, and satisfaction may decrease.

In trials with time delay, the mental workloads of participants were higher than in the no time-delay condition. The reaction times from the stimuli were also slower in the time delay trials. Overall task performance was poorer with system lags as participants' ability to correctly identify targets decreased. Participants tended to easily lose their positional awareness in the

maze, which caused them to incorrectly identify targets as "new" when they had previously identified them.

## Conclusions

Teleoperation requires a strong human-in-the-loop involvement from an operator, and operator control of the robot has a strong impact on overall performance. Previous work in human-computer interaction has demonstrated that human emotions can affect cognitive processes such as human decision making, attention, and working memory (Bechara, Damasio, & Damasio, 2000; Schwarz, 2000; Klein, Moon, & Picard, 2002; Bechara, 2004). While this has been studied in fields such as human-automobile interaction, educational technologies, and health-care systems, there has been comparatively less work in the effect of emotions on telerobotic operators. In this study, we have established a link between the emotional response of frustration and anger to time delays, a common problem in telerobotic control, with the concomitant effects of higher workload, higher physiological arousal, and lower performance. In addition, we have shown that only time delay induced an emotional response, regardless of the task difficulty. Users calibrated their expectations of workload and performance to the apparent task difficulty. Making a task harder does not frustrate them. But when time delay is introduced, they get frustrated/angry at all task difficulty levels.

This examination of teleoperation has studied the implications of time delay not only on the productivity of teleoperation but also the emotional and physical experiences of human operators. Future robotic and computer systems may be able to sense and respond appropriately to the human operators' emotional states in order to achieve a more natural, persuasive, and trustworthy interaction (Fogg 1998; 2002; Voeffray, 2011). The results of this paper identify the

most important features to consider when designing efficacious teleoperation by considering the various aspects of human operators' perceptions.

There are two ways to deal with time delay in systems. One way is to find the cause of the time delay in a system and eliminate it, although this is not applicable to every system. Another approach could be finding the ways to detect operator states and mitigate negative emotions, which could lead to adaptive systems (Scerbo, Freeman, Mikulka, 2003; Feigh, Dorneich, & Hayes, 2012). Adaptive systems that can change behavior to address the sources of negative emotions have the potential to improve task performance, which could lead to higher human-robot joint productivity. Robotic and computer systems that can sense and respond appropriately to human operators' emotional states may enable more natural, persuasive, and trustworthy interactions (Fogg 1998; 2002; Voeffray, 2011). The findings of this study demonstrate the emotional effects of time delay in teleoperation. In future work, we would like to study methods to mitigate the negative emotions of operators. Automation etiquette (Miller & Dorneich, 2006; Miller & Parasuraman, 2007; Dorneich, Ververs, Mathan, Whitlow, & Hayes, 2012) is a promising approach to leverage what is known about human-human etiquette to inform the design of less frustrating HRI. The etiquette in human-machine interaction includes not only politeness but also appropriateness (Hayes & Miller, 2010). The interaction between the interface and the operator could be adapted when the operator meets time delay to mitigate negative emotional responses. Understanding the relationship between time delay and the operators' emotions can provide the foundation for a design that enhances the quality of interactive technologies in teleoperation.

## Acknowledgements

CHAPTER III: EVALUATING HUMAN-AUTOMATION ETIQUETTE STRATEGIES TO
ENHANCE LEARNING

Euijung Yang and Michael C. Dorneich

Abstract

The research investigates how to mitigate user frustration and support student learning through changes in human-computer interaction (HCI) style. Frustration can significantly impact the quality of learning in tutoring. This study examined an approach to mitigate frustration through the use of different etiquette strategies to change the amount of imposition feedback placed on the learner. An experiment was conducted to explore the implications of changing the interaction style between the human and the computer via different etiquette strategies. Participants solved mathematics problems under different frustration conditions with feedback given in different etiquette styles. Changing etiquette strategies led to changes in performance, motivation, confidence, and satisfaction. The most effective etiquette strategies changed when users were frustrated. This work provides the foundation for the design of adaptive tutoring system based on etiquette strategies.

Introduction

Human emotion plays a key role in communication since it can drive the direction of conversation (Ferdig & Mishra, 2004). Previous studies have found that both positive emotions (e.g. happiness and fulfillment) and negative emotions (e.g. frustration and boredom) are key

components in communication, especially in learning (Kort, Reilly, & Picard, 2001; Woolf,

Burleson, Arroyo, Dragon, Cooper, & Picard, 2009; Fisher & Noble, 2009). Negative emotions,

notably frustration, have significant consequences such as lower task productivity (Waterhouse

& Child, 1953; Solkoff, Todd, & Screven, 1964; Klein, Moon, & Picard, 2002; Powers, Rauh,

Henning, Buck, & West, 2011), longer decision making time (Toda, 1980; Bechara, 2004;

Lerner, Li, Valdesolo, & Kassam, 2015), and lower learning efficiency (Graesser, Chipman,

Haynes, & Olney, 2005).

 Human-human tutoring is effective in guiding students through the learning process. It

can support students' learning by responding to questions, analyzing answers, and providing

customized feedback. Intelligent tutoring systems (ITSs) are computer-based instructional

systems whose purpose is to provide customized feedback to users and enable learning in an

effective manner by specifying instructional contents and teaching strategies (Wenger, 1987;

Murray, 2003; Gilbert, Blessing, & Guo, 2015). Research in ITSs attempt to implement the best

methods acquired from traditional tutoring with human tutors and move beyond them to discover

new methods for teaching and learning (Murray, 2003; Broderick, 2011; Koedinger & Tanner,

2013). Systems that include emotion as a factor are called affect-aware systems or affective

systems, and typically provides adaptive feedback and adjust the level of task difficulty of the

problems in order to consider user emotions (Kort, Reilly, & Picard, 2001; Picard et al., 2004).

However, ITSs have limited ability to adjust their interaction behavior based on the emotional

state of the student. This is in contrast to human tutors, which have the ability to adapt their

behavior to appropriately meet the needs of the student (Woolf et al., 2009).

 In human-human interaction, people interact differently when they detect the emotional

states of others (Ekman, 1970; Picard et al., 2004). For example, special communication skills

are used by physicians to deliver bad news when they detect their patients' negative emotions (Back et al., 2007). A human tutor may change his or her speaking style to enhance a student's motivation or mitigate frustration by considering other factors besides performance in order to maximize student learning. Feedback can be used to not only enhance performance, but also to precursors to performance such as motivation, confidence, and satisfaction (Keller, 1987). However, mitigating frustration in human-computer interaction through system changes has been less explored, given the complexity of the interplay between frustration and HCI, and its subtlety (Klein, Moon, & Picard, 2002).

One approach to designing responsive interactions between humans and computers is adaptive systems. Adaptive systems can adjust their behavior by tracking the condition of the users (Feigh, Dorneich, & Hayes, 2012), and have four categories: 1) adapting the allocation of functions between the human and the automation system, 2) adapting the information displayed to the user, 3) changing the user's task priority by directing their attention, and 4) changing the interaction style between the human and the system. Among these four approaches, changing the interaction style is the least explored area due to the interplay of human factors considerations. For instance, while humans use various interaction styles when they face certain situations, adjusting the way computers deliver information violates the human factors principle of consistency in the context of HCI (Feigh, Dorneich, & Hayes, 2012). However, a consistent feedback style may not always be the best in every situation.

Observing the ways to communicate in human-human interaction can provide inspiration to design for HCI. When humans interact with each other, their social behaviors are governed by expectations based on conventional norms between the speaker and hearer. Etiquette is a code of conventional requirements for social behavior. Interactions between people with inappropriate

etiquette may be confusing, unproductive, or even dangerous since people who share the same model of etiquette expect the same level of social behaviors from each other (Wu, et al., 2010). Etiquette has three dimensions: social power, social distance, and imposition. The social power and social distance are decided by the relationship between speakers and hearers. However, the level of imposition can be determined by using different interaction styles since it refers to the amount of demand or burden (Brown and Levinson, 1978; Kasper, 2005). It is possible for people to have expectations when interacting with computers.

The concept of automation etiquette apply human-human etiquette conventions to HCI. (Miller & Funk, 2001; Miller et al., 2004). Once the system has the ability to understand and include user emotions, the interaction between the user and the computer system could be made more sophisticated. Computers could modify their behavior with users in appropriate ways to further joint performance. For instance, in tutoring, human tutors are finely attuned to their students' emotional states. If computers could be more attuned, they may be able to provide appropriate responses in stressful situations where human emotion is impacting the ability to function. Initial studies have been conducted to explore the effects of various interaction styles and the concept of etiquette to potentially enhance human-human tutoring (Pearson, Kreuz, Zwaan, & Graesser, 1995), increase the situation awareness of users in HCI (Wu, Miller, Funk, & Vikili, 2010), and lead to higher reliability of the system from the user's perspective (Parasuraman & Miller, 2004). Advances of tutoring could be combined with human-computer interfaces that incorporate more empathy and affect, enabling technology to move ever closer to authentically embodying the richness of the social interactions between humans (Picard et al., 2004; Woolf et al., 2009).

Understanding the effects of different etiquette strategies on users' performances, motivation, confidence, and satisfaction can contribute to the design of an effective HCI system to enhance the quality of interactions between users and systems. An experiment was conducted to investigate the effects of etiquette strategies in tutoring while the participants solved mathematics problems under different levels of frustration. The goal was to develop an understanding how different etiquette strategies can have differential effects not only performance, but the precursors of motivation, confidence, and satisfaction.

Related Work

Several research areas are relevant in the study of adaptive etiquette strategies. Aspects of user frustration are discussed to understand their impact on performance. Etiquette Strategies are discussed in both human-human interaction and HCI to apply to intelligent tutoring systems. Finally, the role of motivation, satisfaction, and confidence in the learning process will be briefly discussed.

**Frustration and Human-Computer Interaction**

Even though computer systems provide benefits in productivity, frustration is one of the most common experiences in HCI (Ceaparu, Lazar, Bessiere, Robinson, & Shneiderman, 2004). Frustration, defined as an emotional state in which obstacles block the possibility of achieving a goal (Lawson, 1965), is a complex emotion related to anger and disappointment; aggression is one of its consequences (Dollard, Miller, Doob, Mowrer, & Sears, 1939). Frustration has been shown to reduce the quality of ongoing performance by eliciting responses that interfered with the completion of a given task (Waterhouse and Child, 1953). In an experiment conducted on

children, frustration significantly reduced perceptual-motor performance, especially in boys (Solkoff et al., 1964).

Despite the ongoing technological innovations, frustration remains to be a universal problem for users of computers, personal handheld devices, automated systems, or other computer systems. As such, frustration has become a significant interest in the context of HCI. Frustration has been shown to be both frequent and damaging to productivity. Frustration stemming from the use of computers causes users to waste an average of 42-43% of their time (Lazar et al., 2005).

Previous work found that task performance is influenced by the level of frustration. For example, a higher level of frustration led to lower performance score on digit-symbol substitution test (Hokanson & Burgess, 1964). Likewise, operators' task performance was diminished when they were frustrated by system delays in a robot vehicle teleoperating task (Yang & Dorneich, 2015). Frustration led to lower user satisfaction, lower motivation, and drove the users to seek alternative systems (Hoxmeier & DiCesare, 2000; Lazar et al., 2005). In learning, higher frustration caused slower response times (Chen, Gross, Stanton, & Amsel, 1981) and delayed content acquisition (Amsel, 1992). Frustration also reduced the motivation of students (Weiner, 1985), and lead to a lack of confidence of students in computer science (Hansen & Eddy, 2007).

Studies have explored how to account for user frustration in the development of effective tutoring systems. Woolf et al. (2009) used a variety of heuristic strategies to respond to student affect including mirroring student actions to show empathy; adjusting the authority level of the tutoring system to reduce pressure; and changing the voice, motion, and gestures of the avatar in the tutoring system to provide encouragement for the students. The intelligent tutor's strategies

effectively supported the students by encouraging them to continue their tasks although they were frustrated (Woolf et al., 2009; Arroyo et al., 2007). These studies showed that frustration is a topic worth exploring for reasons other than its relation to productivity. Sensing and responding to emotions such as frustration is a step towards building a more human-like affective computer (Picard, 2002).

**Etiquette in Human-Human Interaction**

Etiquette strategies between humans were developed to redress the affronts posed by face-threatening acts (FTAs) (Brown and Levinson, 1978; Mills, 2003). FTAs are an act that inherently damages the face of the addressee or the speaker by acting in opposition to the desires of the other. Positive face is characterized as the desire to be liked, admired, ratified, and related to positively, noting that one would threaten positive face by ignoring someone. Negative face is the desire not to be imposed upon, and to be unimpeded in one's action (Brown and Levinson, 1978). Using etiquette strategies consists of attempting to maintain the hearer's face.

Etiquette can be decomposed into three social variables: social power (i.e., ability of one person to impose their will on another), social distance (e.g. level of familiarity), and imposition (i.e., degree of threat of an FTA). Whilst the social power and the social distance between two entities need long time periods to be changed, if they can be changed at all, the imposition from speaker to hearer can be easily adjusted to mitigate FTAs, thereby forming the basis of different etiquette strategies (Brown and Levinson, 1978).

Etiquette strategies were used to facilitate cooperation to maintain each other's face. There are four types of etiquette strategies: bald, negative politeness, positive politeness, and off-record. A bald strategy is a direct way for a speaker to say something without any consideration to the level of imposition on the hearer. For example, "Pass me the hammer." It does nothing to

minimize threats to the hearer's face. Positive politeness minimizes the social distance and imposition between speaker and hearer by expressing statements of friendship, solidarity, and compliments. For instance, "That is a nice hat, where did you get it?" Negative politeness attempts to be respectful; however, the speaker also assumes that he or she is in some way imposing on the hearer. Examples would be to say, "I don't want to bother you but..." or "I was wondering if..." Off-record utterances use language to give indirect feedback. One says something that is rather general. For example, when the speaker insinuates the listener would turn up the thermostat, saying "Wow, it's getting cold in here." In this case, the hearer must make some inference to recover what was intended in the feedback (Brown & Levinson, 1978).

The concept of etiquette strategies was also employed in tutoring. For example, the effectiveness of different interaction styles with etiquette were examined to see how these strategies could potentially enhance or inhibit effective tutoring (Pearson, Kreuz, Zwaan, & Graesser, 1995). Human tutors were able to select from one of three different etiquette strategies as they saw fit: bald, positive politeness, or negative politeness when they communicated with their students. This study examined how the etiquette strategies were used by human tutors in tutoring conversations, both positively and negatively. Observations from conversation examples of this study show that positive politeness was used to encourage the students when they struggled to solve problems. However, the tutors' responses about the problem answer (e.g., "No, that is wrong.") may lead to negative impressions for students even though it was not the part of intentional feedback based on etiquette strategies. This study suggested that human tutors use different interaction strategies to tailor tutoring even though there were violations of the rules of conversations.

**Etiquette in Human-Computer Interaction**

The concept of etiquette and politeness has been applied to automation (Miller & Funk, 2001; Miller et al., 2004). Miller et al. (2008) developed computational models of communication focused on politeness and etiquette, and established roles of social interactions such as managing power, familiarity relationship, urgency, and indebtedness. Etiquette was used to make natural and polite interactions between humans and computer systems (Parasuraman & Miller, 2004).

Various systems for training and tutoring have explored the concept of etiquette. A virtual manufacturing plant factory training system was developed to teach employees based on two levels of politeness: direct and indirect (polite). Results showed that indirect interaction lead to higher student motivation (Qu, Wang, & Johnson, 2005). The virtual factory training system demonstrated beneficial effects of two etiquette strategies (positive and negative politeness) on learning efficiency (Johnson & Wang, 2010). In a similar manner, a language and culture learning system explicitly delivered language contents and taught social norms by using face-to-face interactions with etiquette and anthropomorphism (Johnson, Friedland, Schrider, Valente, & Sheridan, 2011). A disease and hospital information system was developed to convey information politely (Bickmore, 2010). The participants' ratings of politeness and appropriateness were higher in bald, positive politeness, and negative politeness conditions, but lower in off-record condition because it requires subtlety and consideration of context to be properly comprehended.

**Motivation, Confidence, Satisfaction, and Performance**

In education, various factors influence effective student learning. Keller (1987) proposed four steps for encouraging and sustaining students' motivation in the learning processes:

attention, relevance, confidence, and satisfaction (ARCS). The ARCS model has been used to improve learning effectiveness in distance learning (Malik, 2014), employee education (Visser & Keller, 1990), and manufacturing trainings (Shellnut, Knowltion, & Savage, 1999). Higher levels of motivation, confidence, perceived satisfaction, and overall performance lead to higher rates of engagement in a combination of classroom and online learning (Mohammad & Job, 2012).

## Method

The objective of this study was to explore the ability of etiquette strategies to mitigate user frustration and improve task performance, motivation, confidence, and satisfaction in tutoring.

**Hypothesis**

• H1: Changing etiquette strategies in tutoring lead to differences in performance, motivation, confidence, and satisfaction.

• H2: Participants a priori (baseline) preference of etiquette strategy will not be correlated with the strategy that results in the highest performance, motivation, confidence, and satisfaction under different levels of frustration.

• H3: When users are frustrated, the most effective etiquette strategies are different from when they are not frustrated.

**Participants**

A total 40 university students (23 males, 17 females) averaged 21.1 years old (range: 18 – 29). They averaged 5.7 hours (range: 1 – 15) of computers use daily. Participants' self-assessed math skill levels where measured six subjects on a scale of 0-10: algebra (M=8.65, SD=1.32), geometry (M=7.58, SD=2.02), trigonometry (M=7.35, SD=1.91), calculus (M=8.01, SD=1.70),

statistics (M=6.06, SD=2.40), probability (M=6.14, SD=2.49). Participants last attended

mathematics class an average of 1.35 years ago (range: $1-3$).

**Task**

Participants were asked to solve mathematics problems in algebra, geometry,

trigonometry, calculus, statistics, and probability. Problems were from of the Graduate Record

Examination (GRE) practice books, an exam used for admissions into graduate school. Twenty

problems were provided (see Figure 11), one for each trial. All problems had the same level of

task difficulty (GRE correct rate that $30\% - 40\%$) to ensure that participants would require

feedback frequently in order to solve the problem. Problems were displayed on a computer

monitor with a small stopwatch. Scratch paper and pencils were provided.

*Two trains leave a station at 12:00 pm going in perpendicular directions. If the first train travels 60 mph and the second train travels at 80 mph. At what time is the distance between them exactly 33 and 1/3 miles?*

*Jeff received a 10% increase in his salary in each of the last 3 years. If his present salary is $26,620, what was his starting salary?*

**Figure 11. Example problems.**

**Independent Variables**

The independent variables were *Frustration* (high, low) and *Etiquette Strategy* (bald,

positive politeness, negative politeness, off-record, no feedback).

Frustration was induced by interfering with the ability of a person to attain a goal

(Lawson, 1965). Frustration was elicited by changing the label of the level of task difficulty on

the problems and imposing a time constraint. Even though all problems had the same level of

difficulty, half of the twenty problems that were labeled as 'easy' problems since a mismatch

between expected and the actual level of difficulty can cause frustration (Hone, 2006).

Additionally, a time constraint was also employed to manipulate frustration (Wahlström,

Hagberg, Johnson, Svensson, & Rempel, 2002). Beeps at 1 minute, 30 second and 10 seconds reminded the participant of the time constraint, which was calculated by the average time of five practice problems. The manipulations were designed to elicit frustration without causing the user to simply give up on the task.

Four different etiquette strategies were used to communicate feedback, as well as a no feedback condition as a baseline. Table 4 shows the same feedback being presented in each etiquette strategy.

**Table 4. Example sentences of etiquette strategies.**

| Etiquette Strategies | Definition | Example sentences |
|---|---|---|
| **Bald** | Direct without consideration to level of imposition. | Use appropriate formula. |
| **Positive Politeness** | Minimize imposition via statements of friendship, solidarity, and compliments. | Why don't you try other formulas? Let's check them together! |
| **Negative Politeness** | Respectful but assumes some level of imposition. | If it's alright with you, could you please check other formulas as well? |
| **Off-Record** | Indirect feedback. | Various formulas are provided. |

**Dependent Variables**

*Etiquette Strategies Preference.* The participants were asked before the experiment to rate their preferences for the four etiquette strategies. Participants were asked to read the definitions and examples of four etiquette strategies, and complete their preference rating (on a 10-point Likert scale). This baseline data was employed to compute the correlation between their preference and trial results.

*Independent Variable Manipulation Verification (Frustration).* The independent variable manipulation of frustration was verified via subjective ratings of frustration Participants were asked after every trial their subjective rating of frustration (on a 10-point Likert scale). In addition, the NASA TLX frustration (Hart & Staveland, 1988) subscale scores served as a subjective measure of frustration. To verify the independent variable manipulation, participant responses were compared between low and high frustration in the no feedback condition.

*Task Performance.* A rubric was used to grade their score (see Table 5). TLX

performance subscale scores provided a subjective measure of performance.

**Table 5. Scoring rubric.**

| Score | Answer |
|-------|--------|
| 1 | Correct, variables and equations demonstrated |
| 0.75 | Correct equation with calculation mistakes |
| 0.50 | Correct approach but wrong or no equations |
| 0.25 | Participant defined variables or drew shapes but incorrect approach |
| 0 | Blank |

*Motivation, Confidence, and Satisfaction.* After each trial, participants were asked to rate

motivation, confidence, and satisfaction on a 10-point Likert scale.

*Appropriateness and Effectiveness.* After each trial, participants were asked to rate

feedback appropriateness and effectiveness using Likert scale from 0 – 10.

*Workload.* The participants' mental demand and temporal demand were measured

through NASA TLX subscales after each trial.

**Experimental Design**

This experiment is a 2 (frustration: low vs high) x 5 (etiquette strategy: bald/positive

politeness/negative politeness/off-record/no feedback) within-subject design. Each combination

of independent variables condition was tested twice (20 trials). Condition order was

counterbalanced using Latin squares to account for learning effects.

**Procedure**

The experiment began with the consent process, short briefing, and demographic survey.

Training included a review and practice problems until participants felt comfortable. Completion

time over the last five practice trials were used to set the time constraint for high frustration

trials. Between trials, the participants were asked to complete a post-trial survey and a NASA

TLX. A post-experiment survey gathered opinions and strategies. During debriefing, the

experimenter explained the true goal of the study, as participants were initially told that the study was intended to test their mathematics problem-solving ability. All study materials can be found in Appendix B.

**Data Analysis**

Shapiro-Wilk test was used to check normality of data. Bartlett's test was used to test the homogeneity of variance. Measured data were analyzed with ANOVA tests. Post-hoc analysis used Tukey's test in order to distinguish pairwise means that are significantly different from each other. The results are reported as significant for alpha <.05, and marginally significant for alpha <.10 (Gelman, 2013).Cohen's d was calculated to check an effect size. The Cohen's d results are reported as small effect for .20 < d <.50, medium effect for .50 < d <.80, and large effects for d >.80. Spearman's rank order correlation coefficient was computed to test the association between two ranked variables: participants' baseline preferences of etiquette strategies versus each dependent variable.

**Limitations and Assumptions**

This experiment used only math problems. It is possible that the type of task will greatly influence the optimal feedback strategy. Further work will be needed to generalize the results of this study.

<div align="center">Results</div>

**Interaction Style Preferences**

Before starting the trials, participants' had significantly ($F(3,117)=12.6$, *p<.0001*) different preferences of etiquette strategies. Figure 12a indicates significant pairwise differences

between groups when they do not share a letter the participant preference for a strategy was determined by identifying their highest rank among four strategies (see Figure 12b).



<div align="center">(a)</div>



<div align="center">(b)</div>
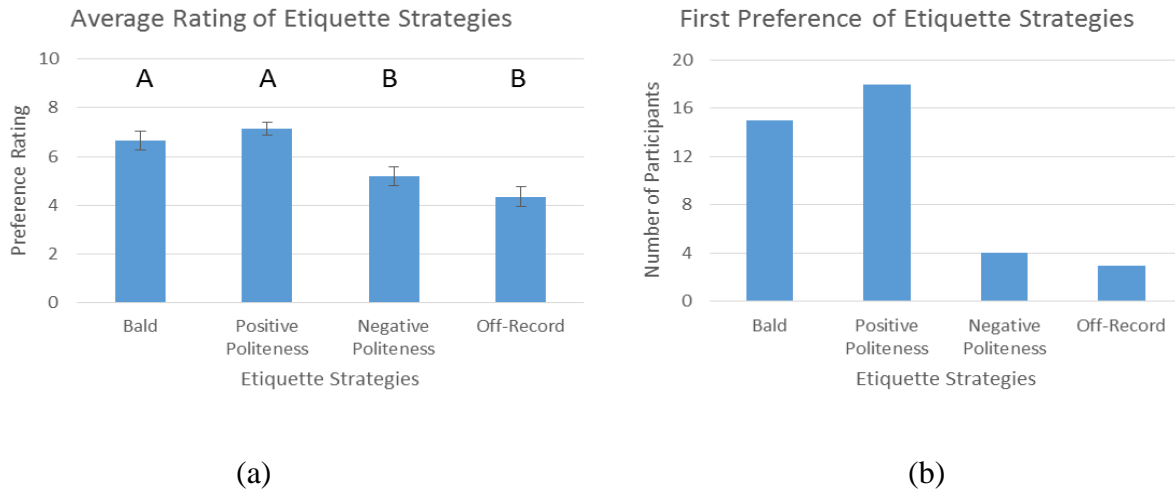
**Figure 12. (a) Average and standard error of strategies preference (n=40). (b) Count of preferred strategy.**

## Independent Variable Manipulation Verification (Frustration)

The TLX frustration subscale was significantly ($F_{(1,39)}=48.53$, *p<.0001*, d=0.72) higher for high frustration than low frustration (see Figure 13a). Participants' subjective rating of frustration showed that the frustration condition was significantly ($F_{(1,39)}=8.31$, *p=.0064*, d=0.56) higher than the low frustration condition (see Figure 13b). The figure indicates significant pairwise differences between groups when they do not share a letter. This verifies the manipulation of frustration though problem labeling and time constraints.
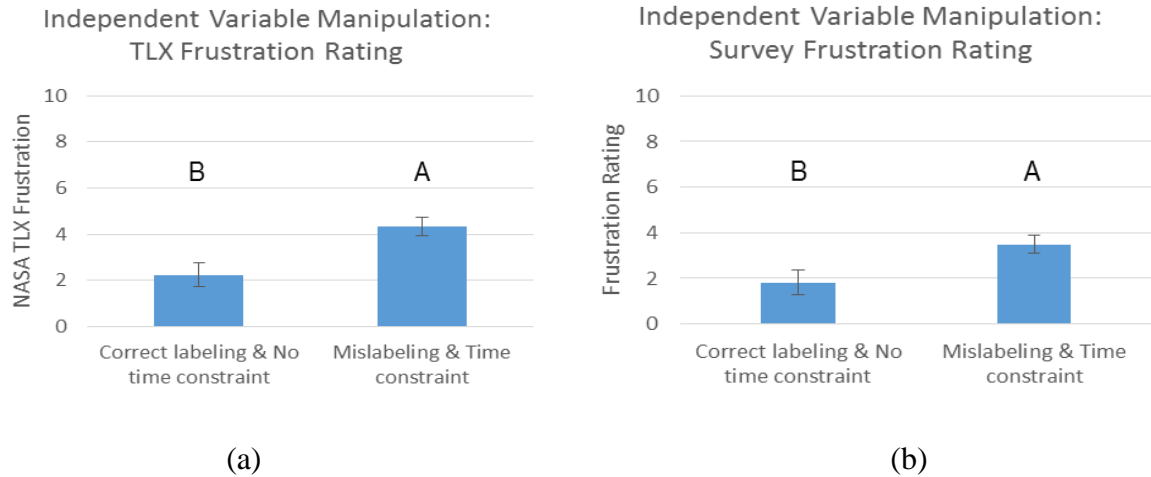
Figure 13. Mean and standard error of frustration (n=40).

Anecdotal participant's comments in the high frustration conditions included: "I do not have enough time to solve problems," "Is it really easy problem?" "I am so frustrated," "There is no hope."

**Task Performance**

The participants correctly solved significantly ($F(1,39)=127.44$, *p<.0001*, d=0.81) more problems in low frustration than high frustration. Etiquette strategies were significant ($F(4,156)=2.77$, *p=.0289*). Figure 14a indicates significant pairwise differences between groups when they do not share a letter. In addition, the interaction effect was also significant ($F(4,156)=3.28$, *p=.0128*).

The participants rated their own performance significantly lower ($F(1,39)=30.24$, *p<.0001*, d=-0.41) in high frustration than low frustration. Etiquette strategies were significant ($F(4,156)=11.64$, *p<.0001*). The interaction was not significant. Figure 14b indicates significant pairwise differences when two groups do not share a letter.
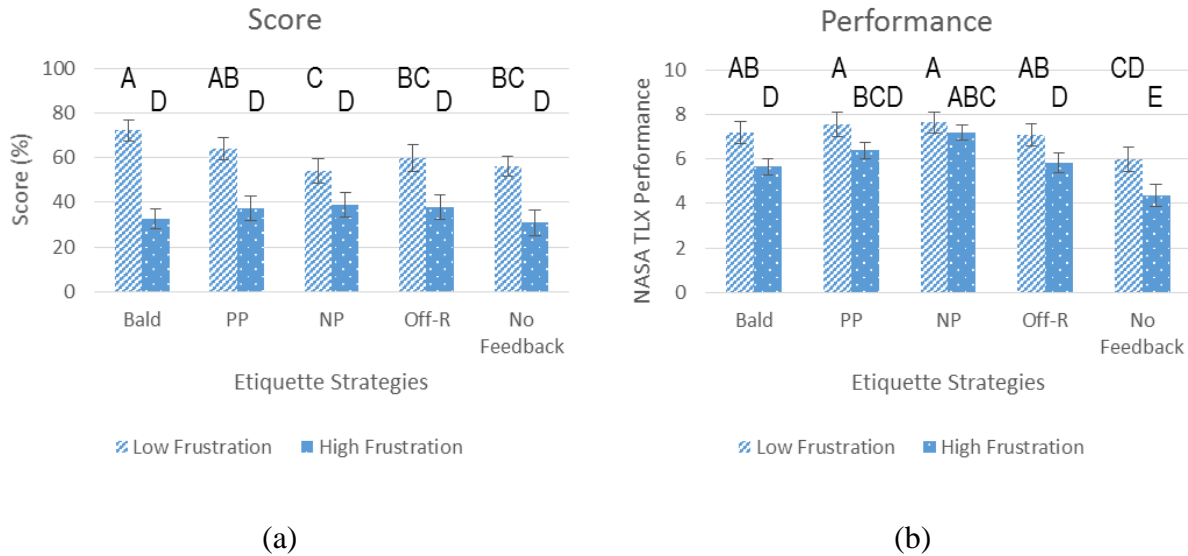
Figure 14. Mean and standard error of (a) problem score and (b) TLX performance (n=40).

There was no correlation between the scores and participants' baseline etiquette strategy preferences. There was no correlation NASA TLX performance rating and participants' baseline interaction etiquette strategy preference.

**Motivation**

Etiquette strategies were significant (F(4,156)=5.45, *p=.0004*). Frustration was not significant. The interaction was not significant. Figure 15a indicates significant pairwise differences between groups when they do not share a letter. There was no correlation between the motivation and participants' baseline etiquette strategy preference.

**Confidence**

Participants had significantly (F(1,39)=12.82, *p=.0009*, d=0.47) more confidence about tasks in low frustration than high frustration. Etiquette strategies were significant (F(4,156)=9.66, *p<.0001*). The interaction was not significant. Figure 15b indicates significant pairwise differences between groups when they do not share a letter. There was no correlation between the confidence and participants' baseline etiquette strategy preference.

(a)                                      (b)
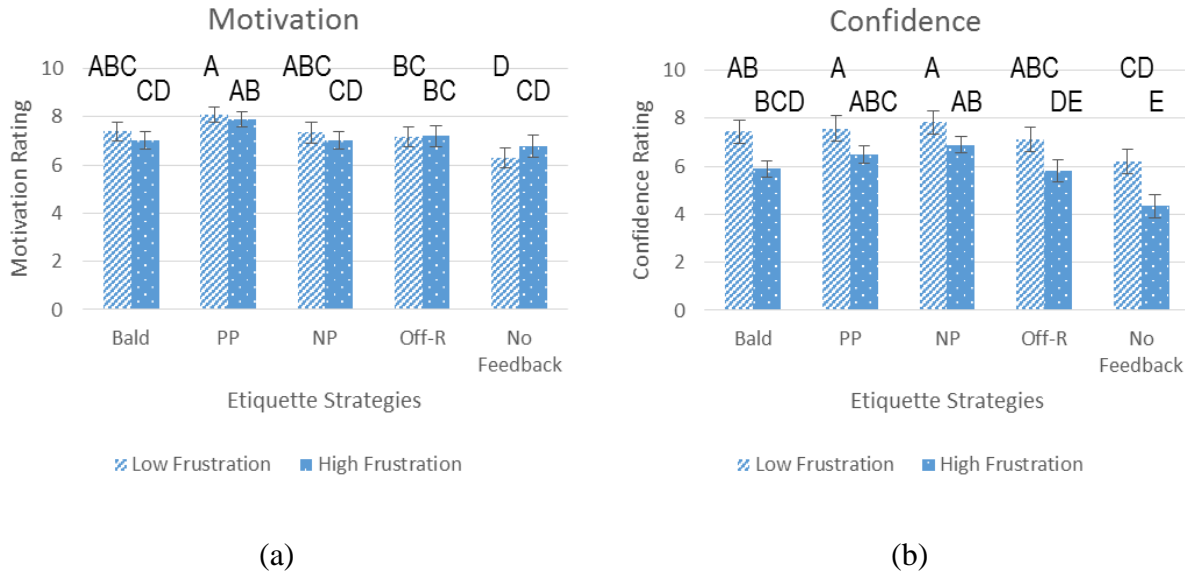
**Figure 15. Mean and standard error of (a) motivation and (b) confidence (n=40).**

## Satisfaction

Participants were significantly (F(1,39)=7.32, *p=.0100*, d=0.22) more satisfied with overall feedback in low frustration than high frustration. Etiquette strategies were significant (F(4,156)=9.43, *p<.0001*). The interaction was not significant. Figure 16a indicates significant pairwise differences between groups when they do not share a letter. There was no correlation between the satisfaction with feedback and participants' baseline etiquette strategy preference.

Participants were significantly (F(1,39)=33.58, *p<.0001*, d=0.31) more satisfied with their own performance in low frustration than high frustration. Etiquette strategies were significant (F(4,156)=10.54, *p<.0001*). The interaction was not significant. Figure 16b indicates significant pairwise differences between groups when they do not share a letter. There was no correlation between the satisfaction with performance and participants' baseline etiquette strategy preference.

(a)  (b)
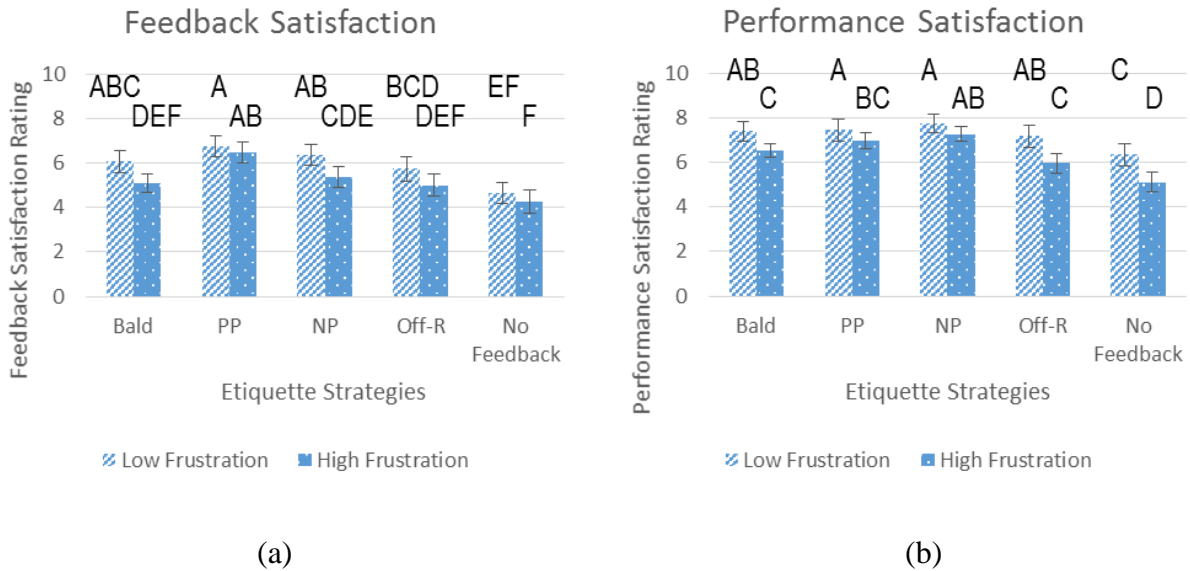
**Figure 16. Mean and standard error of satisfaction with (a) feedback and (b) performance (n=40).**

## Feedback Appropriateness and Effectiveness

*Feedback Appropriateness.* Etiquette strategies were significant ($F(4,156)=12.31$, $p<.0001$). Frustration was not significant. The interaction was not significant. Figure 17a indicates significant pairwise differences between groups when they do not share a letter. Feedback appropriateness and participants' baseline etiquette strategy preference was significantly correlated ($r=-0.11$, $p=.0323$).

*Feedback Effectiveness.* Feedback was marginally significantly ($F(1,39)=3.06$, $p=.0878$, $d=0.14$) more effective in low frustration than high frustration. Etiquette strategies were significant ($F(4,156)=10.31$, $p<.0001$). The interaction was not significant. Figure 17b indicates significant pairwise differences between groups when they do not share a letter. There was no correlation between the feedback effectiveness and participants' baseline etiquette strategy preference.
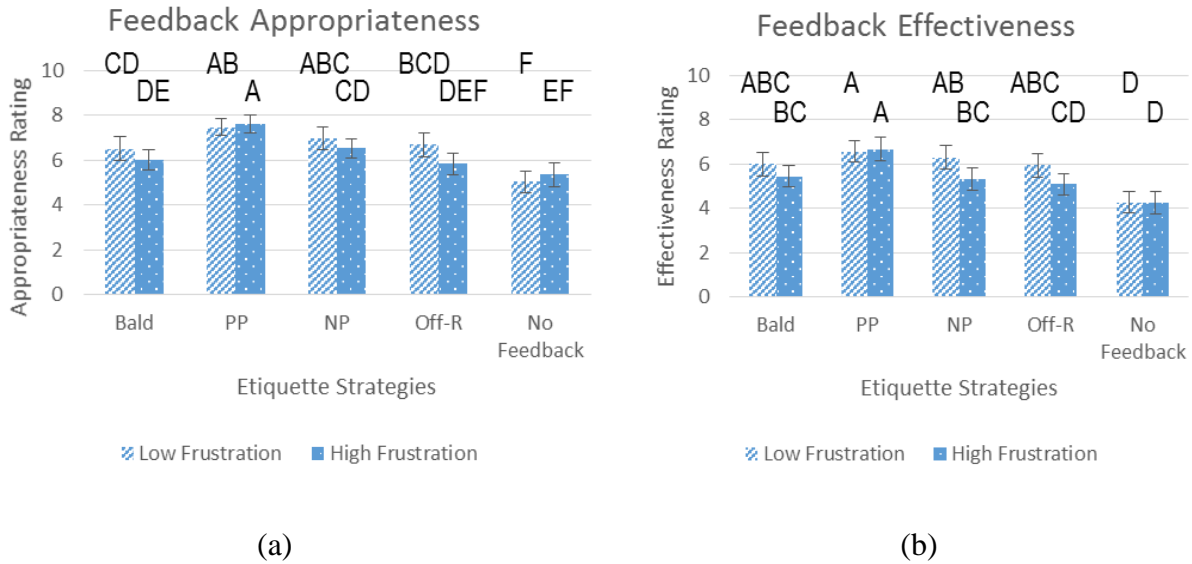
Figure 17. Mean and standard error of feedback (a) appropriateness and (b) effectiveness (n=40).

## Workload

*Mental Demand.* Etiquette strategies were significant ($F(4,156)=6.69$, *p<.0001*). Frustration was not significant. The interaction was not significant, Figure 18a indicates significant pairwise differences between groups when they do not share a letter. There was no correlation between the mental demand and participants' baseline etiquette strategy preference.

*Temporal Demand.* Feedback was significantly ($F(1,39)=70.25$, *p<.0001*, d=1.23) more temporally demanding in low frustration than high frustration. Etiquette strategies were significant ($F(4,156)=4.82$, *p=.0011*). The interaction was not significant. Figure 18b indicates significant pairwise differences between groups when they do not share a letter. There was no correlation between the temporal demand and participants' baseline etiquette strategy preference.
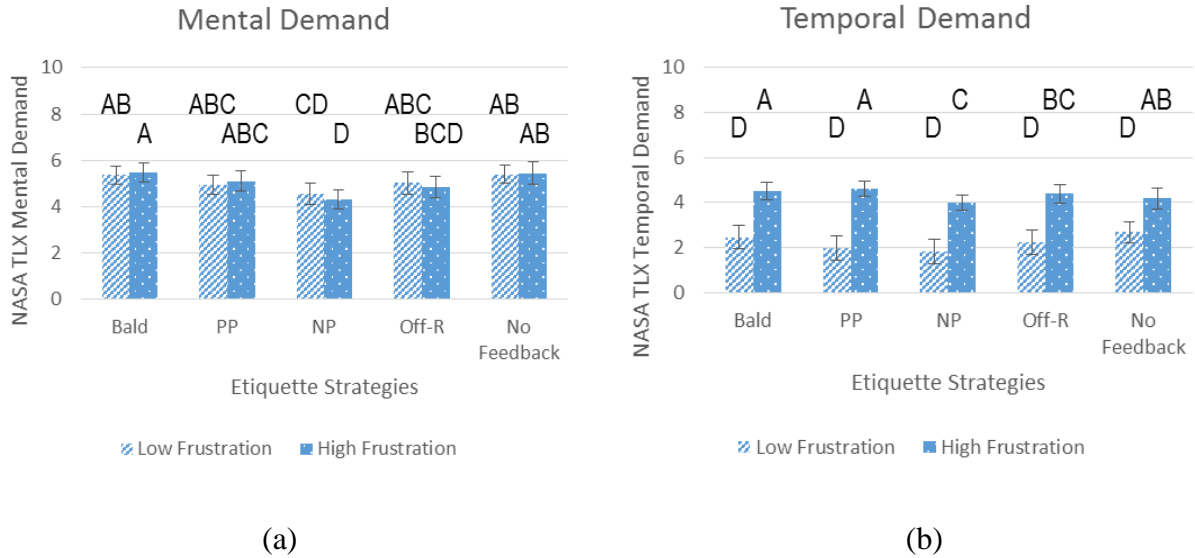
Figure 18. Mean and standard error of TLX (a) mental demand and (b) temporal demand (n=40).

## Discussion

The results of the study showed that etiquette strategies significantly influence performance, motivation, confidence, and satisfaction. The first hypothesis stated: "Changing etiquette strategies in tutoring lead to changes in performance, motivation, confidence, and satisfaction." This hypothesis was partially supported. Scores from the mathematics problems were higher when the bald strategy was provided in low frustration condition whilst there were no differences in the scores between any etiquette strategies in high frustration condition. However, the scores from high frustration condition may have been affected by a ceiling effect due to the time constraint. Some participants did not have enough time to solve the given problems. Negative politeness leads to higher performance than positive politeness in high frustration condition.

Positive politeness led to higher motivation and satisfaction than no feedback in the low frustration condition. On the other hand, motivation and satisfaction were not driven by

interaction style of the feedback in the high frustration condition. People who were provided feedback with negative politeness showed higher confidence about their own work than people who were not given any feedback in high frustration condition. Moreover, positive politeness led to higher satisfaction with feedback than no feedback in high frustration condition. Thus negative politeness and positive politeness effectively worked to increase confidence and satisfaction with feedback. These results demonstrated that user's performance, motivation, confidence, and satisfaction vary depending upon etiquette strategies in tutoring.

The second hypothesis stated: "Participants a priori (baseline) preference of etiquette strategy will not be correlated with the strategy that results in the highest performance, motivation, confidence, and satisfaction." This hypothesis was supported. There was no correlation between four dependent variables and participants' baseline etiquette strategy preferences. These results provide evidence that people's performance, motivation, confidence, and satisfaction can be affected by a change of etiquette strategy, and is not fixed based on their own preferences. Thus it may be feasible to build an adaptive tutoring system that changed interaction styles to improve to performance, motivation, confidence, and satisfaction.

The third hypothesis stated: "When users are frustrated, the most effective etiquette strategies are different from when they are not frustrated." This hypothesis was fully supported. When participants were frustrated and provided feedback with positive and negative politeness, their self-assessed performance, motivation, confidence, and satisfaction were higher than when they were provided bald, off-record, and no feedback. Thus, the most effective etiquette strategies were different when users are frustrated.

The most beneficial etiquette strategy to use at a given moment depends on the users' emotional state (e.g. frustration level) and the current state of their learning (as measured by

confidence, satisfaction, motivation, and performance). Based on the results of this study, a set of

rules were developed to determine which etiquette strategy to utilize under different conditions.

Chapter IV details how this rule set was derived. This rule set will be used to implement an

adaptive tutoring system (as described in Chapter V).

Although frustration is a common and natural emotion people experience while learning,

it has impacts on learners' self-esteem, distractibility, and ability to follow directions (Liu,

Pataranutaporn, Ocumpaugh, & Baker, 2013). A tutor's feedback can be a great help to mitigate

students' frustration and ultimately reduce the consequences of frustration. The results of this

study show that different feedback interaction styles impact different aspects of the learning

process. For example, the participants performed better by receiving feedback based on bald and

positive politeness under low frustration while they performed better with negative politeness

feedback under high frustration. Their satisfaction with performance showed a similar pattern:

participants were more satisfied when they received positive politeness feedback under low

frustration, but negative politeness feedback under high frustration. These results demonstrated

that different etiquette strategies were helpful to improve the participants' performances when

they were highly frustrated. It provides the evidence that proper interaction style can mitigate the

influences of frustration. Likewise, the participants' ratings of motivation, satisfaction, and

confidence showed a similar tendency. Since motivation, satisfaction, and confidence are directly

connected to the students' learning goals, providing appropriate feedback to support these is

crucial to enhance effective learning (Keller, 1987). These results can be applicable for not only

a human tutor but also a computer tutor.

## Conclusion

Results showed that providing feedback with a different interaction style based on user emotional states can vary the results of tutoring. However, it does not mean that one strategy was obviously better across all four dependent variables. It indicates that different strategies impact the dependent variables in different ways. Further studies need to establish the interaction of strategy influences. The results of this study can be used to guide the HCI between an intelligent tutoring system and a student in the domain of math tutoring.

Frustration is one of the most frequently occurring emotions in situations where people use computers (Ceaparu et al., 2004) and learn (Woolf et al., 2009). Understanding user states and adapting the behavior of the system could be used to mitigate frustration, improve interactions between the human and the system, and potentially improve task performance. Adaptive systems could mitigate user frustration by changing the way it communicates with users, much in the same way a human tutor would change his or her feedback when a student becomes frustrated. This study provided a basic understanding of the role of different interaction styles of feedback under varying user emotional states, and can be used to form the basis of an adaptive tutoring system.

In human-computer tutoring, most of the real-time adaptation is triggered by poor performance and results in a change to the task difficulty. However, a good human tutor will be aware of the emotional state of the learner and adapt their interaction style to support aspects of the student's learning that underlie performance such as a student's motivation, confidence, or satisfaction. Future work will look at the ability to adapt interaction styles depending on the emotional state of the students as well as the goal of the tutor. These results which derive the logic of etiquette strategies will form the basis of an adaptive tutoring agent.

Acknowledgements

CHAPTER IV: DEVELOPMENT OF A SET OF RULES TO ADAPT ETIQUETTE

STRATEGIES

A set of rules was developed to determine which etiquette strategy to utilize under different conditions based on the results of Chapter III. This chapter detailed the approach taken to develop the rule set.

In Chapter III, the level of confidence, satisfaction, motivation, and performance was measured after each math problem. The level of frustration and the etiquette strategy used when provide feedback changed after each problem in a syntactic exploration of two levels of frustration and the four etiquette strategies. Statistical analysis (ANOVA and post-hoc analysis suing Tukey's test) were conducted in order to distinguish pairwise means that were significantly different from each other. Based on the Tukey's test, the highest performing etiquette strategy or strategies (denoted in Tukey's results as strategies that included letter A) were selected as the most effective etiquette strategies for each learning factor and frustration level combination. Sometimes a single strategy was chosen and sometimes a group of strategies was chosen. Specifically, the following strategies were chosen based on the results of Study 2:

- *Confidence*. In high frustration, positive politeness and negative politeness were picked as the best strategies because they were equally likely to improve performance (they all included the letter A). Bald, positive politeness, and negative politeness were selected in low frustration because they were in the top group that included letter A.

- *Satisfaction*. In high frustration, positive politeness and negative politeness were picked as the best strategies. Positive politeness was rated as the highest strategy based on the feedback satisfaction rating and negative politeness was rated as the highest strategy

based on the performance satisfaction rating. Positive politeness was selected in low

frustration because it was in the top group that included letter A based on both the

feedback and the performance satisfaction rating.

- *Motivation*. For the motivation in high and low frustration, positive politeness was picked

  as the best strategies because it was the highest rated group that included letter A.

- *Performance*. In low frustration, bald and positive politeness were selected because they

  were in the top group that included letter A based on the score results. However, due to

  the ceiling effect on the score results, they could not be used to select a strategy for the

  high frustration condition. Thus, in high frustration, the result from NASA TLX

  performance rating was used since the score in high frustration. Negative politeness was

  picked as the best strategy because it was the top group that included letter A.

These results show that which etiquette strategy is the most supportive depends on what

learning factor is targeted and the level of frustration (see Table 6).

**Table 6. Highest performing etiquette strategies for each combination of frustration and learning factor.**

|  | Confidence | Satisfaction | Motivation | Performance |
|---|---|---|---|---|
| **High frustration** | Positive Politeness, Negative Politeness | Positive Politeness, Negative Politeness | Positive Politeness | Negative Politeness |
| **Low frustration** | Bald, Positive Politeness, Negative Politeness | Positive Politeness | Positive Politeness | Bald, Positive Politeness |

After establishing the best etiquette strategies for each learning factor, a logic of rule set

was developed. The rule set provides a systematic method for determining what learning factor

should be targeted next. First, the system checks if any of the learning factors' ratings are at or

below 5 (on a 10-point scale), and targets the factor with the lowest score. If two ratings share

the same lowest score, the choice is determined by considering the flowing priority order (as

determined in part by the ARCS model): confidence, satisfaction, motivation, performance. The

ARCS places confidence before satisfaction. Once students are satisfied and confident, their

level of motivation increases (Mohammad & Job, 2012), placing it third in the sequence. Finally,

students' performance was the fourth factor in the sequence because it increased when students

were confident, satisfied, and motivated with their learning (Keller, 1987; Mohammad & Job,

2012).

If all learning factors ratings are above 5, then the system determined which factor had

the largest decrease from the previous measurement. Again, ties are broken by the same order as

above. Once the target learning factor has been identified, the strategy chosen is listed in Table 6.

Some combinations of learning factor and frustration level have more than one possible etiquette

strategies. Thus, if a strategy fails to improve the targeted factor, then the strategy chosen the

next time the factor if targeted changes, as shown in Table 6. This is based on Study 2, which

demonstrated that several strategies may improve a particular learning factor, so the system was

designed to cycle between them until the most effective one is found. This is the first step

towards individualization the adaptations through learning. As an example, if the targeted

learning factor was confidence, for low frustration, the first strategy tried will be bald. If that

fails to improve confidence, the next time confidence is targeted, the strategy chosen will be

positive politeness. The logic for the rule set is summarized in Table 7.

**Table 7. Feedback logic for a prototype.**

| | Feedback Logic |
|---|---|
| 1 | If any learning factors' ratings <= 5 (on 10-point scale), target factor with lowest score |
| 2 | If all learning factors' ratings > 5, then choose factor with largest decrease from previous measurement |
| 3 | If tie, priority order: motivation, confidence, satisfaction, performance |
| 4 | If strategy fails to improve targeted factor, then choose next alternative strategy (if applicable) |

The rule set is used to determine the etiquette strategies after measuring the learning

factors and the level of frustration. After gathering all ratings of the learning factors and the

frustration level, the system uses a rule set to determine how to modify its feedback with

etiquette strategies. This process is conducted after each problem (see Figure 19). This rule set

will be used to implement an adaptive tutoring system (as described in Chapter V).
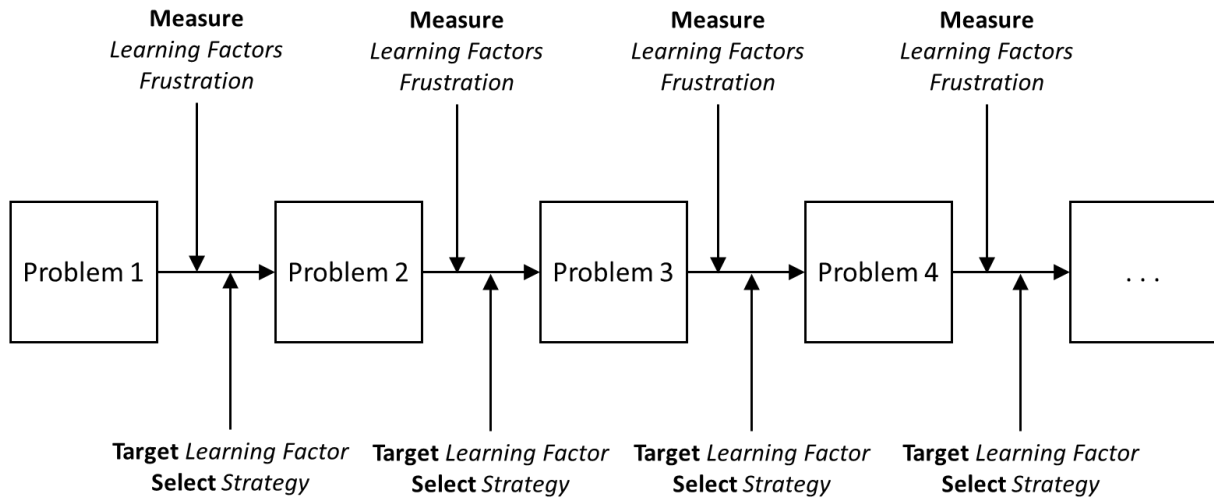
**Figure 19. Application of rule set to an adaptive tutoring system.**

CHAPTER V: EVALUATING ADAPTIVE TUTORING SYSTEM BASED ON HUMAN-
AUTOMATION ETIQUETTE STRATEGIES

Euijung Yang and Michael C. Dorneich

Abstract

The purpose of this research is to evaluate an adaptive tutoring system based on etiquette strategies in the context of human-computer interaction (HCI). This study investigated the adaptive interaction styles of an intelligent tutoring system that was based on etiquette strategies. Previous work demonstrated that different feedback etiquette strategies have differential effects on students' motivation, confidence, satisfaction, and performance. The best etiquette strategy is also determined by the level of user frustration. Based on these findings, an adaptive tutoring system prototype was developed where the tutor feedback's etiquette style changes dynamically to best address students' needs. An experiment was conducted to explore whether the selection of proper etiquette strategies for a given situation has implications for learning. Participants solved mathematics problems under different frustration conditions with feedback that had different adaptation levels of etiquette strategies. The results demonstrated that feedback with systematic adaptation increase motivation, confidence, satisfaction, and performance when it targets to improve. It shows that proper interaction style can mitigate the influences of frustration.

## Introduction

Feedback is an essential part of effective learning and helps to improve students' achievements; it is used to provide encouragement, praise, critique, and evaluation. Feedback advises students on how to approach, analyze, and learn in their efforts to reach their goals (Bransford, Brown, & Cocking, 2000). Proper feedback improves students' understanding of the subject-matter's contents and provides guidance to enhance their learning (Hattie & Timperley, 2007). Additionally, students' confidence, self-awareness, and enthusiasm can be increased when they receive appropriate feedback during learning (Bellon, Bellon, & Blank, 1992). Studies have shown that that providing more interactive feedback can be better than teaching more content because substantial feedback leads to clear conceptual understanding of the subject (Hattie, 2008; Marzano, Pickering, & Pollock, 2001). A function of feedback is to support the students when they are solving problems. To provide proper feedback when the students need it, tutors should have an awareness of each stage in the students' problem-solving process (Gordon & Bruch, 1974; Dickman & Gordon, 1985).

One of the ways that human tutors know their feedback works is by checking their students' performance (e.g., test score, grades). However, evaluating feedback effectiveness by using only performance is not sufficient since various elements influence learning (Keller, 1987; Mohammad & Job, 2012). Thus, another approach to examining feedback effectiveness is by assessing not only the students' performance, but also motivation, confidence, and satisfaction. The students' motivation, confidence, satisfaction, and performance can be used as an insight into the students' learning because higher levels of these attributes lead to higher engagement (Mohammad & Job, 2012). The attention, relevance, confidence, satisfaction (ARCS) model has been proposed as a suitable method to investigate effective feedback (Keller, 1987). The ARCS

model is a systematic problem-solving approach that requires an understanding of human motivation as part of the learning process and has been used to improve the students' engagement and learning effectiveness. These methods promote the assessment of learning effectiveness because they enable the students to engage learning contents, improve problem-solving skills, and interpret the role of feedback in learning.

Person to person tutoring can be effective because tutors provide appropriate and timely feedback throughout the learning process; he or she is able to support students' learning by responding to questions, analyzing answers, and encouraging students. Intelligent tutoring systems (ITSs) are computer-based instructional systems that provide feedback to users and enable learning in an effective manner by specifying instructional content and teaching strategies (Wenger, 1987; Murray, 2003; Gilbert, Blessing, & Guo, 2015). ITSs attempt to implement the best methods acquired from traditional human-human tutoring and move beyond it to discover new strategies for teaching and learning (Murray, 2003; Broderick, 2011; Koedinger & Tanner, 2013). However, ITSs with the ability to adjust their behavior based on the student's states – such as motivation, confidence, and satisfaction – have been less deeply examined. This is in contrast to human tutors, which have the ability to adapt their behavior to appropriately meet the needs of the student (Woolf, Burleson, Arroyo, Dragon, Cooper, & Picard, 2009).

Previous studies mainly took into account task performance as a trigger to adapt tutors' feedback (Wood & Wood, 1999; Zakharov, Mitrovic, & Ohlsson, 2005; Roll, Aleven, McLaren, & Koedinger, 2011). Yet this approach does not consider other aspects that are crucial in learning: students' motivation, confidence, and satisfaction. Furthermore, feedback may be different depending on what the system is trying to accomplish. Thus, if the system had methods for detecting a variety of user states, the system could vary the way it delivers feedback to users.

Customized feedback could produce better learning experiences because users' conditions (e.g., emotion, motivation) are also crucial factors in learning and HCI (Klein, Moon, & Picard, 2002).

Human tutors are keenly aware of more than the students' performance, but also their motivation, confidence, and satisfaction. Sometimes human tutors adapt just to help with those attributes, even if the students' performance is only a byproduct. In addition, human tutors may change their feedback to help students cope with negative emotions, notably frustration, which can have significant consequences such as lower task productivity (Waterhouse & Child, 1953; Solkoff, Todd, & Screven, 1964; Klein, Moon, & Picard, 2002; Powers, Rauh, Henning, Buck, & West, 2011), longer decision making time (Bechara, 2004; Lerner, Li, Valdesolo, & Kassam, 2015), and lower learning efficiency (Graesser, Chipman, Haynes, & Olney, 2005). Yet it is not easy for computer tutors to adjust their feedback depending on their current assessments of the situation in real-time. If the goal is to design a computer tutor with the ability to take into account students' frustration, motivation, confidence, and satisfaction, then work needs to be done to design feedback strategies that positively impact emotions and motivation, beyond simply adapting the task difficulty when the students perform badly.

Observing the ways humans communicate can provide inspiration to the design of ITSs. When humans interact with each other, their social behaviors are governed by expectations depending on conventional norms. Etiquette is a code of requirements for social behaviors that has been applied into design considerations in HCI (Miller & Funk, 2001; Miller, Wu, & Funk, 2008). Communication without etiquette may cause confusing, unproductive, or even dangerous situations since people who share the same mental model of etiquette expect the same level of social behaviors from each other (Wu, Miller, Funk, & Vikili, 2010). Etiquette has three dimensions: social power, social distance, and imposition. Social power and social distance are

decided by the relationship between speakers and hearers. However, the level of imposition can be determined by using different interaction styles since it refers to the amount of demand or burden of the hearers (Brown and Levinson, 1978; Kasper, 2004). Etiquette strategies enable changing the level of the imposition of communication (Brown and Levinson, 1978).

One approach to designing effective feedback is changing the interaction style of the ITSs depending on students' condition along multiple dimensions. Etiquette strategies may provide the basis of an adaptive interaction style of feedback by adjusting the level of imposition between tutors and students. Students' frustration may be mitigated if a computer tutor is able to decrease the level of the imposition of the students at appropriate times. Thus, selecting proper etiquette strategies for a given situation may be able to enhance students' learning by diminishing their negative experiences. Preliminary work has suggested that etiquette strategies differential effects on students' motivation, confidence, satisfaction, and performance in certain situations (Yang & Dorneich, 2016). If computer tutors could be more finely attuned to their students' conditions, by observing and detecting the students' states, computer tutors may be able to provide the proper responses in stressful situations where the human states impact their ability to function.

In this paper, a tutoring system was developed that provided adaptive feedback depending on students' states. Students' motivation, confidence, satisfaction, and performance were used as the triggers for adaptation. In addition, the effect of frustration was also investigated to scrutinize how adaptation works in stressful situations. The next section discusses the related work to understand background studies. A description of adaptive tutoring system prototype, the experimental method, and experiment's results are described in subsequent sections to investigate the effects of adapting etiquette strategies based on student state.

Discussion and conclusion provide detailed reflections from the results and future work

suggestions.

Related Work

Various aspects of feedback and learning are discussed to understand their impacts on the

students. The concept of adaptive systems is also reviewed as the approach towards providing

adaptive feedback in a tutoring system. Definitions and applications of ITSs and affect-aware

systems are reviewed to take into account students' emotions in HCI. The notion of etiquette and

its strategies are discussed in both human-human interaction and HCI to apply to an adaptive

tutoring system.

**Feedback and Learning**

Feedback is an indispensable element of education, it facilitates growth in a student's

potential at different stages in learning. It also enables the students to become aware of their

strengths, weaknesses, and which actions they need to employ to improve achievement (Hattie &

Timperley, 2007; Norcini, 2010). Studies developed ways to provide effective feedback to

enhance students' learning. For example, Wiggins (2012) suggested seven keys to increase

feedback effectiveness: goal-referenced, tangible and transparent, actionable, student-friendly,

timely, ongoing, and consistent. In a similar manner, methods to provide effective feedback were

developed: specific, realistic, concentrated on student behavior, balanced content, and continued

support (Dempsey, 1993; London, 2003; McGill & Brockbank, 2003).

Analyzing processes of learning and problem solving have been advantageous to foster

productive feedback. A tutor's awareness of a student's problem-solving process is crucial

because the tutor frequently interferes in the student's problem-solving to provide timely

feedback. Guiding students through specified problem-solving steps allows the tutors to teach students the ways to approach and solve a given problem (Gordon & Bruch, 1974; Dickman & Gordon, 1985). Six steps of the problem-solving process were established to categorize the students' learning process: identifying the problem or situation, brainstorming solutions or generating alternatives, evaluating the alternative solutions or suggestions, deciding on final solutions, implementing the solution or decision, and conducting a follow-up evaluation. Those steps have been used to develop effective feedback and provide proper rationales behind of generating feedback content (Dickman & Gordon, 1985; Gordon, 2008).

While analyzing learning processes, students' perceptions are crucial components (Keller, 1987; Keller & Kopp, 1987). The ARCS model was developed to investigate effective ways of interpreting students' perceptions during learning. The ARCS model is a systematic approach, which has four elements for encouraging and sustaining students' motivation in the learning processes: attention, relevance, confidence, and satisfaction (Keller, 1987; Keller & Suzuki, 1988; Keller, 2009).

Pedagogical studies explored the effectiveness of the ARCS model. For instance, the impact of the ARCS model's ability to overcome the lack of motivation was examined by applying it to distance learning students. The ARCS model approach increased students' attention during instruction, established relevant feedback to their needs, generated a positive impression for their confidence, and provided satisfying experiences by emphasizing their achievements. After providing feedback based on the ARCS model, student's motivation to learn was improved (Malik, 2014). For employees who attended staff development classes, their motivation to learn was increased by receiving techniques based on the ARCS model, such as various supporting materials, stories relevant to the learning contents, motivational messages,

and compliments on their learning attitudes (Visser & Keller, 1990). Higher levels of motivation, confidence, perceived satisfaction, and overall performance lead to higher rates of student engagement. These results demonstrated that those four factors enable tutors to provide encouragement and promote students' effective learning (Mohammad & Job, 2012). Similarly, an online feedback system in virtual environments leads to higher levels of motivation, satisfaction, and performance when compared to no-feedback (Geister, Konradt, & Hertel, 2006).

**Adaptive Systems**

Adaptive systems have an ability to automatically tailor their behavior to best support human performances (Feigh, Dorneich, & Hayes, 2012). Adaptive systems can change their behavior in four broad categories: modification of function allocation, modification of task scheduling, modification of interaction, and modification of content. Among these four categories, the modification of interaction contains how information is delivered between the human and the system. While previous studies focused on determining the point at which feedback is provided, the style by which it is provided has been less utilized in adaptive systems due to the view that changing interaction style violates consistency from the human factors perspective (Feigh, Dorneich, & Hayes, 2012).

A representative example of adaptive systems is the CAT, which selects the next test problem from a bank of questions depending on the system's current assessment of the student's ability level. (Sands, Waters, & McBride, 1997; Linacre, 2000). If the students successfully solve the first problem, the next one will be more difficult, and vice versa (Wainer et al., 2000). Traditional fixed testing produces the best estimates for students who have intermediate abilities and less accurate estimates for students who have advanced or novice abilities. However, the CAT estimates students' ability level and adjusts its level of difficulty accordingly (Wainer et al.,

2000). The CAT saves time because it does not attempt to provide problems that are too difficult or too easy for the students (Sands, Waters, & McBride, 1997; Van der Linden & Glas, 2000).

**Intelligent Tutoring and Affect-Aware Systems**

Human-human tutoring is effective in guiding students through the learning process. The tutor can support students' learning by responding to questions, analyzing answers, and providing customized feedback. Computer systems, called ITSs, were developed to try to replicate the benefits of a human tutor by providing feedback to learners. ITSs help them by providing feedback including hints, supportive materials, and problem-solving method suggestions (Wenger, 1987; Murray, 2003; Koedinger & Tanner, 2013). Traditionally, students' performance was the main factor of ITSs' feedback development (Ong & Ramachandran, 2003; Goldberg, Holden, Brawner, & Sottilare, 2011). However, recent studies have started to include students' emotions as factors within ITSs because emotions influence attention, problem-solving, and working memory in learning (Sylwester, 1994). For example, a companion learning system was implemented to provide adaptive feedback; it adjusts the level of task difficulty of the problems to consider the user's negative emotions (e.g., frustration, boredom, depression). In other words, the students received easier problems when they experience negative emotions (Kort, Reilly, & Picard, 2001; D'Mello et al., 2008).

These type of systems are called affect-aware systems. Affect is used to indicate the experience of feeling or emotion in psychology (Martin, Hogg, & Abrams, 2010); it is also considered as an important factor in personal and social life (Izard, Kagan, & Zajonc, 1984). Affect-awareness is an ability to recognize what an individual is feeling (e.g., happy, sad, frustrated) and manage the situation depending on the individual's emotions (Kaliouby, Picard, & Baron-Cohen, 2006). Affect-aware systems are any kind of systems that contains an ability to

consider a user's emotions as an element of the system (D'Mello et al., 2008; Woolf et al., 2009). These are implemented to deliver different types of feedback including encouragement, empathy, and mirroring emotions of the users based on their emotional states (Picard et al., 2004; Picard, 2006; Woolf et al., 2009; Calvo & D'Mello, 2012). Affect-aware systems have the ability to adjust its behavior to consider the users' emotions beyond their performance when they experience negative emotions (Kort, Reilly, & Picard, 2001).

There are a number of studies that investigated the roles and consequences of affect-aware systems in various fields. Affect-aware systems strive to improve the performance and satisfaction of an individual by considering his or her affect states. For instance, when users were provided affect-support (e.g., providing a vent button to allow users express their emotions, concerns, and problems) while frustrated due to from feedback delay, they played the computer game longer, had lower frustration, and higher satisfaction than those who did not provide affect support (Klein, Moon, & Picard, 2002). Another example of affect-aware systems improving user performance is Koko, it has a logic that detects user emotions and provides appraisal feedback when the users feel frustrated or sad. The users who were supported by the Koko performed better than those who did not have the Koko (Sollenberger, & Singh, 2012).

ITSs have been developed to provide computer-aided instruction to support student learning (Kort & Reilly, 2002; Salman, 2013). ITSs typically interact with learners through dialogues, graphical user interfaces, or a rich simulation of the task domain. ITSs also interpret users' inputs and generate the feedback to the users (Anderson & Koedinger, 1997; Granić, Stankov, & Glavinić, 2000; Kort & Reilly, 2002; Padayachee, 2002; Corbett & Anderson, 2008; Nkambou, Mizoguchi, & Bourdeau, 2010; Salman, 2013). 'Emotionally intelligent tutoring systems (EITSs)' were developed to provide adaptive feedback to users by gathering users' self-

assessed emotional states (Ochs & Frasson, 2004). Observation of facial expressions, gross body language, and the content and tone of speech can be gathered to interpret users' emotional states (Kort & Reilly, 2002).

In terms of HCI, various affect-aware systems were developed to adjust the systems' behaviors, based on user emotions, to increase learning efficiency. Smart Tutor, a web-based adaptive interactive tutoring system, changed the content of information based on objectives, skills, knowledge, and emotions of the users (Gamalel-Din, 2002). A similar approach was taken by MetaTutor, which provided realistic verbal dialogue to determine when and if to provide feedback based on users' frustration, surprise, and happiness. The real-time prompt feedback reduced the level of negative emotions of the users compared to no-feedback (Azevedo, Witherspoon, Chauncey, Burkett, & Fike, 2009; Harley, Bouchet, Hussain, Azevedo, & Calvo, 2015). A recent review (Calvo & D'Mello, 2012) of research activities in affect-aware systems and their learning applications evaluated how students attest to different emotional states such as boredom, confusion, engagement, and curiosity during a tutoring session. They concluded that emotion detection is a developing research area and is expected to be integrated into future generations of ITSs to improve learning outcomes by responding to the user's affective states.

**Etiquette Strategies**

Etiquette strategies mitigate or soften direct expressions of desire or intention between humans. Goffman (1967) defined 'face' as the self-image that people want to claim as themselves, thus human face could be an element of etiquette within human society. Brown and Levinson (1987) developed a model of politeness called face-threatening acts (FTAs) which show strategies of politeness from individual's self-esteem (face). Etiquette strategies were developed to redress the affronts to the face posed by FTAs to addresses. These are also the ways

for the speakers to mitigate face threats carried by certain FTAs to hearers (Brown & Levinson, 1978; Mills, 2003).

The concept of FTAs was based on both positive and negative face. Positive face refers to one's self-esteem while negative face refers to one's freedom to act (Foley, 1997; Miller, Wu, and Funk, 2008). Brown and Levinson (1978) characterized positive face as the desires to be liked, admired, ratified, and related to positively, noting that one would threaten positive face by ignoring someone. Negative face as the desire not to be imposed upon, and to be unimpeded in one's action (Brown and Levinson, 1978). Using etiquette strategies consists of attempting to keep the hearer's face. An act of face threatening is defined as an act that inherently damage the face of the addressee or the speaker by acting in opposition to the wants and desires of the other.

Etiquette strategies were used to facilitate cooperation which is needed amongst the speakers and hearers to maintain each other's faces since the positive and negative faces are the basic desires in any social interaction (Foley, 1997; Brown & Levinson, 1978). Brown and Levinson (1978) identified four types of etiquette strategies: *bald, negative politeness, positive politeness,* and *off-record*. A bald strategy is a direct way for a speaker to say something without any consideration to the level of imposition on the hearer. For example, "Pass me the hammer." It does nothing to minimize threats to the hearer's face. Positive politeness minimizes the social distance between speaker and hearer by expressing statements of friendship, solidarity, and compliments. For instance, "That is a nice hat, where did you get it?" Negative politeness attempts to be respectful; however, the speaker also assumes that he or she is in some way imposing on the hearer. Examples would be to say, "I don't want to bother you but..." or "I was wondering if..." Off-record utterances use language to give indirect feedback. One says something that is rather general. For example, when the speaker insinuates the listener would

turn up the thermostat, saying "Wow, it's getting cold in here." In this case, the hearer must make some inference to recover what was intended in the feedback (Brown & Levinson, 1978).

There have been studies of the linguistic aspects of etiquette within the human-human interaction. For example, Kasper (2004) investigated linguistic etiquette and identified variables of it from data-based studies such as social power (i.e., containing an interlocutor's relative positions in social hierarchies, age, gender, and language impairment), social distance (i.e., politeness appears to be expended in negotiable relationships with familiars, but decreases within both intimates and strangers), and imposition (i.e., including requesting, urgency, apologizing, thanking, indebtedness, and complaining). Whilst the social power and the social distance between two people only change slowly over time, the imposition from speaker to hearer can be easily adjusted to mitigate FTAs, thereby forming the basis of different etiquette strategies (Brown & Levinson, 1978; Kasper, 2004). Another example of the linguistic approach of etiquette is Grice's (1975) Maxims, which provide efficient ways for conversation within the human-human interaction. These maxims of conversation are 1) the maxim of quality (be truthful); 2) maxim of quantity (be brief); 3) maxim of relation (be relevant to the topic); and 3) maxim of manner (be clear and avoid ambiguity). These maxims underlie the etiquette of interactions between people.

**Etiquette Strategies and Learning**

The concept of etiquette strategies was also employed in tutoring. For example, the effectiveness of different interaction styles with etiquette based on Grice's (1975) maxims was examined to see how these strategies could potentially enhance or inhibit effective tutoring (Pearson, Kreuz, Zwaan, & Graesser, 1995). Pearson et al. (1995) provided a five-step dialogue frame to the human tutors who taught mathematics and let them communicate with their students

by following the dialogue frame. The human tutors were able to select from one of three different etiquette strategies as they saw fit: bald, positive politeness, negative politeness when they communicated with their students. They examined how the etiquette strategies were used by human tutors in conversations with students, both positively and negatively. Observations from conversation examples showed that positive politeness was used to encourage the students when they struggled to solve problems. However, the tutors' responses about the problem answer (e.g., "No, that is wrong.") may lead to negative impressions for students even though it was not the part of intentional feedback based on etiquette strategies. This study suggested that etiquette strategies could be used in various ways to enhance or inhibit tutoring even though there were violations of the rules of real conversations (Pearson et al., 1995).

In a preliminary work (Yang & Dorneich, 2016), upon which this current study is based, the level of imposition in a learning context was varied to see if different etiquette strategies had differential effects under different conditions. Specifically, etiquette strategies were used to change the level of imposition and mitigate students' frustration. Changing etiquette strategies led to changes in performance, motivation, confidence, and satisfaction. However, the previous study used etiquette strategies in isolation without consideration of the student's current state. Each strategies impact on motivation, satisfaction, confidence, and performance was studied, under different levels of student frustration. The current study build upon these results to develop a rule set that specifies under which conditions (triggers) different strategies are used to provide feedback during the students' problem-solving process. The goal is to create a tutoring system where the feedback's etiquette style changes dynamically to best address students' needs.

Adaptive Tutoring System Prototype

A prototype tutoring system was developed to test the effectiveness of dynamically adapting the human-computer interaction to improve student motivation, confidence, satisfaction, and performance. The adaptive tutoring system prototype was designed to provide feedback with etiquette strategies while the participants solve mathematics problems. The system tutored on college-level mathematics problems.

## Tutoring Domain

The adaptive tutor system provided mathematics problems in six different subjects (i.e., algebra, geometry, trigonometry, calculus, statistics, and probability). All problems were from practice books of the Graduate Record Examination (GRE). Figure 20 shows example problems.

*At a neighborhood gathering, the ratio of children to adults is four times the ratio of adult men to adult women. If there were twice as many adult women as adult men and 10 men showed up, how many children were present?*

*Oil is poured on a flat surface, and it spreads out forming a circle. The area of this circle is increasing at a constant rate of $5 cm^2/s$. At what rate, in cm/s, is the radius of the circle increasing when the radius is 5 cm?*

**Figure 20. Example task problems.**

## Design Rationale

The ARCS model of the motivational design includes a sequence among the learning elements. Once the students start to pay attention during instruction, the tutors need to have relevant learning contents, then the students earn the confidence of what they learn, then the students are satisfied with their learning process, and finally the students are motivated to learn more (Keller, 1987). This study took into account the order of the ARCS model without attention and relevance because the participants were asked to pay attention for the experiment and recruited from science, technology, engineering and mathematics (STEM) fields that are already relevant to the mathematics task itself. While using the sequence of the ARCS model, this

experiment included motivation and performance instead of attention and relevance because students' motivation, confidence, satisfaction, and performance were used to indicate learning effectiveness (Mohammad & Job, 2012). Previous work (Yang & Dorneich, 2016) had demonstrated that those learning factors were increased by receiving feedback with certain etiquette strategies. For instance, bald led to higher performance, positive politeness led to higher motivation and satisfaction, and negative politeness led to higher confidence and satisfaction. However, the most effective strategies sometimes changed if the user was frustrated.

**Feedback Design**

Gordon's (2008) six steps of the problem-solving process were used to develop feedback content. Each step had one to three feedback comments including at least one proactive (e.g., "Define the variables.") and at least one reactive (e.g., "It's not the appropriate formula.") comment. This feedback followed the steps of students' problem-solving process to implement effective feedback and provide justification for feedback contents. Each comment was also designed by applying Wiggins' (2012) seven keys of effective feedback design. For instance, bald feedback was actionable and transparent (e.g., "Recheck what you've done and thought of other possibilities."), positive and negative politeness feedback is student-friendly (e.g., "I think there's something missing. Let's check all the possibilities together!"), and so on. In addition, all comments focused on student behavior and balanced contents. The feedback was specific to each problem and differed between problems. All feedback information was given via as recorded human voices played from the computer speaker.

**Feedback Logic**

In this experiment, a rule set was developed to trigger the most appropriate etiquette strategy as the basis for systematic adaptation. After each math problem, the system decides

which factor needs the most support by measuring the level of motivation, confidence, satisfaction, and performance. Then the system uses a rule set to determine how to modify its feedback with etiquette strategies to one of four possibilities: bald, positive politeness, negative politeness, and off-record (see Table 8).

**Table 8. Example sentences of etiquette strategies.**

| Etiquette Strategies | Example sentences |
|---|---|
| Bald | Use appropriate formula. |
| Positive Politeness | Why don't you try other formulas? Let's check them together! |
| Negative Politeness | If it's alright with you, could you please check other formulas as well? |
| Off-Record | Various formulas are provided. |

The rule set provides a systematic method for determining what learning factor should be targeted next. First, the system checks if any of the learning factors' ratings are at or below 5 (on a 10-point scale), and targets the factor with the lowest score. If two ratings share the same lowest score, the choice is determined by considering the flowing priority order (as determined in part by the ARCS model): motivation, confidence, satisfaction, performance. If all learning factors ratings are above 5, then the system determined which factor had the largest decrease from the previous measurement. Again, ties are broken by the same order as above. Once the target learning factor has been identified, the strategy chosen is listed in Table 9. Furthermore, if a strategy fails to improve the targeted factor, then the strategy chosen the next time the factor if targeted changes, as shown in Table 9. This is based on Study 2, which demonstrated that several strategies may improve a particular learning factor, so the system was designed to cycle between them until the most effective one is found. This is the first step towards individualization the adaptations through learning. As an example, if the targeted learning factor was confidence, for low frustration, the first strategy tried will be bald. If that fails to improve confidence, the next time confidence is targeted, the strategy chosen will be positive politeness.

**Table 9. Rule set of etiquette strategies.**

| Learning Factor | Low Frustration | High Frustration |
|---|---|---|
| Motivation | Positive Politeness | Positive Politeness |
| Confidence | 1) Bald, 2) Positive Politeness, 3) Negative Politeness | 1) Positive Politeness, 2) Negative Politeness |
| Satisfaction | Positive Politeness | 1) Positive Politeness, 2) Negative Politeness |
| Performance | 1) Bald, 2) Positive Politeness | Negative Politeness |

The most effective etiquette strategies differed when users were frustrated. Bald led to higher performance, positive politeness led to higher motivation and satisfaction, and negative politeness led to higher confidence and satisfaction when the students were less frustrated. Negative politeness led to higher performance and confidence, positive politeness led to higher satisfaction when the student was highly frustrated. However, motivation was not driven by interaction style of the feedback when the students were highly frustrated. These results demonstrated that students' performance, motivation, confidence, and satisfaction vary depending upon etiquette strategies in tutoring (Yang & Dorneich, 2016).

The rule set checked three aspects by using the participants' responses: the absolute numbers of those ratings, differences between previous and current trial ratings, and order of dependent variables. The absolute numbers of ratings meant that if any factors among motivation, confidence, satisfaction, and performance were rated lower than 5 out of 10, the feedback was changed to address that factor. If all the factors were rated more than 5, the rule set calculated the differences between previous and current trial ratings. The rule set picked the factor that had the largest decrease and provided the feedback with suitable etiquette strategies for that factor. Finally, based on the ARCS model, the rule set prioritized the following order: motivation, confidence, satisfaction, and performance. Students' performance was the last factor considered because it increased when students were motivated, confident, and satisfied with their learning (Keller, 1987; Mohammad & Job, 2012).

**User Interface**

  The tutoring system prototype was designed to provide real-time feedback while the students solved mathematics problems (see Figure 21 The prototype had six menus on the left side including the types of mathematics subjects. When the students clicked one of the menus, it had four buttons upper side of the screen that indicated four problems. Some problems were labeled easy and some hard. The mathematics question was occurred in the middle of the screen. The feedback is provided via recorded voice messages while the students were solving the problems.
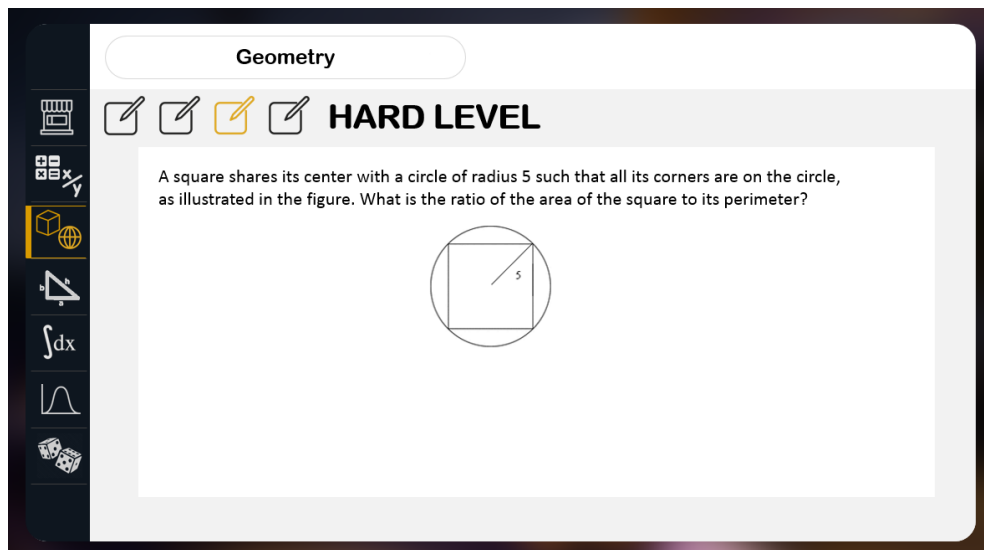


**Figure 21. The screenshot of the adaptive tutoring system prototype.**

**Adaptation Implementation**

  The feedback is provided in one of the four etiquette strategies. The current adaptive tutoring system prototype is tested using a Wizard-of-Oz method: during the trial when the student is solving math problems, the experimenter observed the progress and decided when to trigger feedback. A guide was developed with specific errors or misconceptions for the experimenter to look for, and the appropriate feedback to trigger. Through this method, the

experimenters could ensure that all feedback was triggered correctly, and any effects uncovered in the study would be solely due to the etiquette strategies. After each math problem is completed, the student rated their motivation, confidence, satisfaction, and the experimenter calculated their score (performance). These numbers were used to quickly determine if the feedback strategy should be changed for the next trial (math problem).

## Method

### Objective

The objective of this study is to d investigate the effect of dynamically adapting the interaction style during a tutoring session to best support students motivation, confidence, satisfaction, and performance. The effectiveness of adapting etiquette strategies will be tested by comparing a systematic adaptation based on the rules set (described in the previous chapter) versus a random adaptation. Furthermore, they will be tested in high and low frustration situations.

### Hypothesis

There are two hypotheses to fulfill the research objectives.

• H1: Systematic adaptation of etiquette strategies increases motivation, confidence, satisfaction, and performance more than random etiquette strategies.

• H2: Systematic adaptation of etiquette strategies mitigates user frustration more than random etiquette strategies.

### Participants

A total 31 university students (18 males, 13 females) participated in the experiment. The participants' average age was 25.48 (range: 19 – 31). All subjects were experienced computer

users who used computers on average 8.01 hours per day (range: 3 – 16). Also, all participants had a normal or corrected-to-normal vision in order to exclude the possibility of diminished attention due to vision problems. All participants were from science, technology, engineering, and mathematics fields. Participants' self-assessed math skill levels were measured six subjects on a scale of 0-10: algebra (M=8.15, SD=1.48), geometry (M=7.04, SD=1.89), trigonometry (M=6.75, SD=1.82), calculus (M=7.82, SD=1.79), statistics (M=6.63, SD=2.46), probability (M=6.90, SD=2.24). Participants last attended mathematics class an average of 2.58 years ago (range: 1 – 5).

**Task**

Participants were asked to solve mathematics problems with an adaptive tutoring system. All math problems were from practice books of the Graduate Record Examination (GRE). All problems were chosen from a similar difficulty level, where the historical GRE correct answer rate was 30% – 40%. Scratch paper and pencils were provided.

**Independent Variables**

The two independent variables were *Frustration* (low, high) and *Adaptation* (systematic, random). Frustration can be induced by interfering with the ability of a person to attain a goal (Lawson, 1965; Dollard, Miller, Doob, Mowrer, & Sears, 1939). Frustration was elicited by changing the information of the level of task difficulty on the problems. Even though all problems had the same level of difficulty, half of the twenty problems that were labeled as 'easy' problems to induce frustration. Recognizing a difference between the level of difficulty and the expectation of the difficulty level can cause frustration (Hone, 2006). By labeling a hard problem as easy, a person has unrealistic expectations of the ease of attaining the goal. Additionally, a social pressure also employed to manipulate frustration (Wallace, 1979; Liu & Yu, 2011;

Powers, Rauh, Henning, Buck, & West, 2011). The participants were asked to achieve a high score, which was almost unattainable, and they received false information that indicated all previous participants earned the unattainable score (Powers et al., 2011). The frustration manipulation was controlled to produce enough frustration to affect learning but not too high to cause the user to simply give up on the task.

The adaptation was manipulated by adjusting how the feedback is triggered: systematic and random etiquette strategies. The systematic condition used the rule set (described in the previous chapter) to determine if the strategy should change after a math problem. The random condition provided feedback to the participants by randomly choosing one of etiquette strategies regardless of the results of the previous math problem. The systematic condition was compared to a random condition rather than a no-feedback baseline since any feedback, no matter how poorly designed, could have a benefit from incidental cognitive engagement by simply varying the usual workflow.

Other adaptation manipulations using the same math problems and same feedback were studied in Yang & Dorneich (in review): no feedback, and static feedback for each trial. In that study (see Chapter III), the feedback was tested (in isolation) against a no-feedback condition and found to benefit the learning process, and thus "no feedback" condition was not needed for this study. In addition, no-feedback was also not used in this experiment because the system wanted to provide same amount of the information through feedback instead of providing less amount of the information through no-feedback. Similarly, static adaptation that includes only one etiquette strategy (based on each participant's priori preference) showed that there was no correlation between participants' priori preferences and dependent variables. Thus the current study focused on systematic versus random adaptation comparisons.

**Dependent Variables**

The dependent variables were *task performance, motivation, confidence,* and *satisfaction.* In addition, *frustration* was measured to verify that the independent variable manipulation of low and high frustration was effective. Distribution of strategies and number of strategy transitions were counted to check how many times strategies changed. Moreover, target influences and tradeoffs were calculated to evaluate whether the adaptation impact on the level of motivation, confidence, satisfaction, and performance. Finally, *feedback appropriateness, cognitive workload, emotional states,* and *stress* were also measured. The dependent variables are described in Table 10.

**Table 10. The measurements for both independent variables verification and dependent variables.**

| Variables | Metric | Measurement (Unit) | Frequency | Data Type |
|---|---|---|---|---|
| Independent Variable Manipulation Verification (Frustration) | TLX Subscale Frustration | Scale 0 – 10 | After each math problem within trial | Subjective |
| | Frustration Questionnaire | Likert Scale 0 – 10 | After each math problem within trial | Subjective |
| Task Performance | Problem Solving Score | Score 0 – 10 | After each math problem within trial | Objective |
| Motivation | Motivation Questionnaire | Likert Scale 0 – 10 | After each math problem within trial | Subjective |
| Confidence | Confidence Questionnaire | Likert Scale 0 – 10 | After each math problem within trial | Subjective |
| Satisfaction | Satisfaction Questionnaire | Likert Scale 0 – 10 | After each math problem within trial | Subjective |
| Distribution of Strategies and Number of Strategy Transitions | Number of Strategy Transitions Count | 0 – 5 (times) | After each math problem within trial | Objective |
| Effectiveness when Targeting an improvement in a learning factor | Change in level of Motivation, Confidence, Satisfaction, and Performance between math problems | (-10) – (+10) | After each math problem within trial | Objective |
| Feedback Appropriateness | Appropriateness Questionnaire | Likert Scale 0 – 10 | After each math problem within trial | Subjective |
| Cognitive Workload | TLX Subscale Mental Demand | Scale 0 – 10 | After each math problem within trial | Subjective |
| Stress | Stress Questionnaire | Likert Scale 0 – 10 | After each math problem within trial | Subjective |

*Frustration.* Frustration was an independent variable manipulated by both mislabeling the problems and providing time constraint. Participants were asked after every problem their subjective rating of frustration (on a 10-point Likert scale). In addition, the NASA TLX frustration (Hart & Staveland, 1988) subscale scores served as a subjective measure of frustration. In order to verify that the independent variables manipulation was successful, participant responses were compared for the low and high frustration in the no feedback condition.

*Task Performance.* Task performance was the score from mathematics problems, based on the rubric in Table 11.

**Table 11. The rubric to score problems.**

| Score | Answer Sheet |
|---|---|
| 10 | Answer is correct with variables and equations demonstrated |
| 7.5 | Answer has correct equation but having calculation mistakes |
| 5.0 | Answer has correct approach to solving such as setting up the variable from problems or drawing shapes based on given problems but having wrong equations or no equations |
| 2.5 | If the participant tried to make variables or draw shapes but they were not correct approach |
| 0 | If the answer sheet doesn't have anything |

*Motivation, Confidence, and Satisfaction.* After each math problem, participants were asked to rate motivation, confidence, and satisfaction on a 10-point Likert scale.

*Distribution of Strategies and Number of Strategy Transitions.* Both the Systematic and random adaptation trials changed its feedback depending on motivation, confidence, satisfaction, and performance. Those factors were applied after the first problem, which means that the feedback strategy may change between problems 2 – 5.

*Effectiveness when Targeting an Improvement in a Learning Factor.* After each math problem, the system decides which element needs to be supported by measuring the level of motivation, confidence, satisfaction, and performance. The change in these constructs was then

measured by calculating the difference between two consecutive problems. Thus the effectiveness and the tradeoffs of the mitigation can be assessed for the targeted and non-targeted factors. To answer the question, *"Was the targeted learning factor improved by the system?"*, a 2x2 ANOVA analysis will determine the effect of adaptation and frustration on the targeted learning factor. To answer the question, *"What also happened to the non-targeted learning factors?"*, two 2x4 ANOVA (one for systematic adaptation and one for random adaptation) were conducted with frustration and learning factor as variables.

*Feedback Appropriateness.* After each problem, participants were asked to rate feedback appropriateness using Likert scale from 0 – 10.

*Cognitive Workload.* The participants' mental demand was measured through the NASA TLX mental demand subscale after each problem.

*Stress*. After each problem, participants were asked to their stress using Likert scale from 0 – 10.

**Experimental Design**

This experiment is a 2 (frustration: low vs. high) x 2 (adaptation: systematic vs. random) repeated measures, within-subject design. A within-subject design was used to block the effect of individual differences such as level of skill in solving the math problems. Although switching between high and low frustration may have lessened the overall difference between the two levels of frustration manipulation, it was determined through pilot tests that the within-subject design still resulted in sufficient separation between low and high frustration to test the hypotheses.

There were four trials, two in each combination of intendent variables. Each trial included 5 problems, where the etiquette strategies varied between problems. The order of each

combination of the two independent variables was counterbalanced across participants using a

series of 4x4 Latin squares to account for any learning effects (see Table 12).

**Table 12. Counterbalanced Trials.**

| Trial | Participant 1 | | Participant 2 | | Participant 3 | | ... |
|---|---|---|---|---|---|---|---|
| | Frustration | Adaptation | Frustration | Adaptation | Frustration | Adaptation | |
| Trial 1 | High | Systematic | Low | Systematic | High | Random | ... |
| Trial 2 | Low | Random | High | Systematic | High | Systematic | ... |
| Trial 3 | Low | Systematic | Low | Random | Low | Systematic | |
| Trial 4 | High | Random | High | Random | Low | Random | |

**Procedure**

The experiment began with the consent process, a short briefing, and a demographics

survey. Participants were provided refresher training on mathematics problems and solved

practice problems until they felt comfortable. All study materials can be found in Appendix D.

Participants completed four trials. They were able to have a break whenever they wanted.

Within a trial, after every math problem, participants rated their motivation, confidence, and

satisfaction. Between problems, the participants were asked to complete a post-problem survey

and a NASA TLX. After finishing all four trials, the participants were asked to fill out a post-

experiment survey to gather their opinions, strategies, and ideas to improve future studies.

During the debriefing, the experimenter explained to the participants that the true goal of the

study was to study the effect of etiquette strategies, as they had been initially told that the study

was intended to test their mathematics problem-solving ability. The experiment used a Dell

Precision T1700 desktop PC as a testing apparatus.

**Data Analysis**

The Shapiro-Wilk test was used to check normality of data. Bartlett's test was used to test

the homogeneity of variance. Measured data were analyzed with ANOVA tests. The results are

reported as highly significant for a significance level alpha $<.001$, significant for alpha $<.05$, and

marginally significant for alpha <.10 (Gelman, 2013). Post-hoc comparisons were cealcuated

using Tukeys HSD for pairwise comparisons between conditions. Additionally, a Cohen's d was

calculated to check an effect size which provides a standard measure that expresses the mean

difference between two groups in standard deviation units. The Cohen's d results are reported as

small for .20 < d <.50, medium for .50 < d <.80, and large for d >.80.

**Limitations and Assumptions**

This experiment used only one type of task – math problems. It is possible that the type

of task will greatly influence the optimal feedback strategy. Further work will be needed to

generalize the results of this study to account for this possibility.

Since the task consists of mathematics problems from GRE, it requires an ability to solve

certain level of mathematics problems. The degree of such problem-solving ability, however,

varies from individual to individual and might have influenced the task outcome. Nevertheless,

one of the inclusion criteria of this experiment was recruiting particular people who are majoring

in science, technology, engineering, and math education to mitigate this limitation.

Results

**Frustration – Independent Variable Manipulation Verification**

The TLX frustration subscale was significantly ($F(1,30)=91.1$, *p<.0001*, d=0.87) higher

for high frustration than low frustration (see Figure 22a). Participants' subjective rating of

frustration showed that the frustration condition was significantly ($F(1,30)=19.96$, *p=.0032*,

d=0.59) higher than the low frustration condition (see Figure 22b). The figure indicates

significant pairwise differences between groups when they do not share a letter. This verifies that

the manipulation of frustration was effective though problem labeling and time constraints.
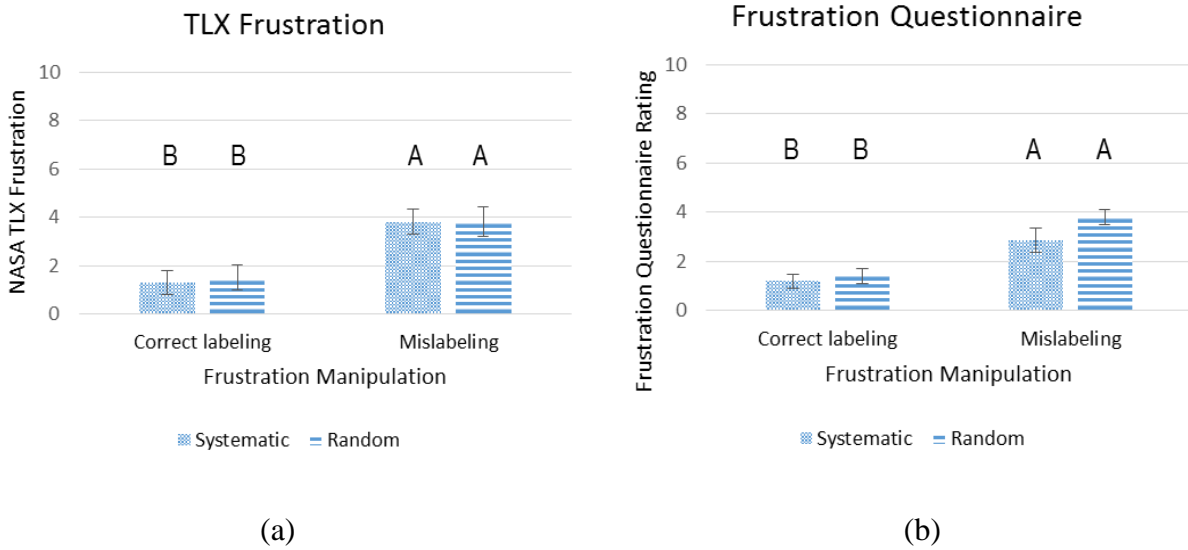
**Figure 22. Mean and standard error of frustration (n=31).**

## Distribution of Strategies and Number of Strategy Transitions

Figure 23 shows the distribution of feedback strategies for each problem in the 5-problem trial sequence in systematic adaptation (a) and random adaptation (b) conditions. The average number of feedback type transition in systematic adaptation condition was 2.45 out of 5 and in random adaptation condition was 4 out of 5. This demonstrates the tutoring system was changing the interaction styles between problems in each trial.
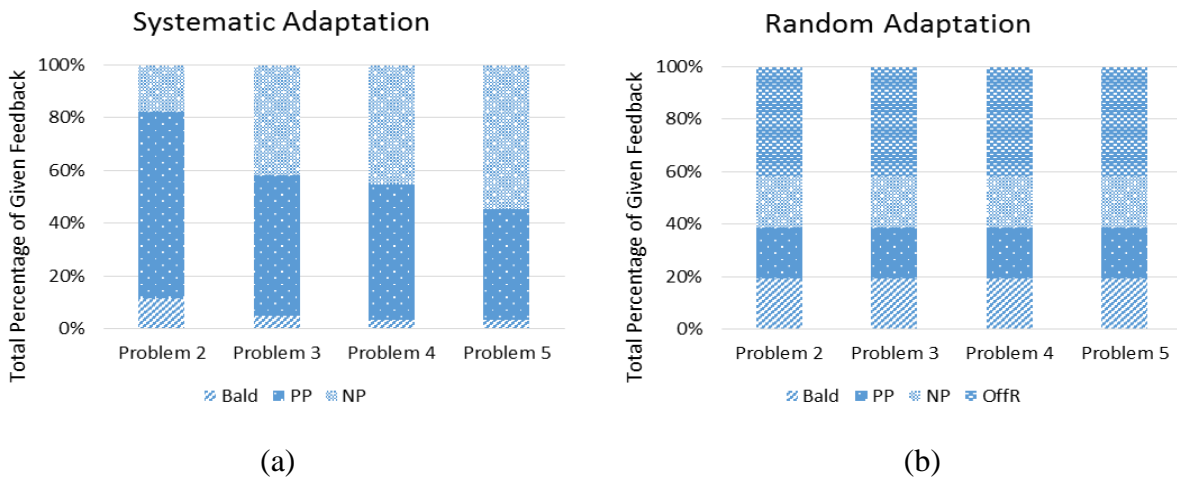


**Figure 23. Distribution of given feedback types in (a) low frustration condition and (b) high frustration condition (n=31).**

**Effectiveness when Targeting an Improvement in Learning Factor**

*Targeting Motivation.* In order to compare the effect of adaptation and frustration, a 2x2 ANOVA was conducted. Systematic adaptation resulted in a significantly ($F(1,11)=11.2$, $p=.0065$, d=0.51) larger gain in motivation than the random adaptation (see Figure 24). However, the effect of frustration, and the interaction between adaptation and frustration were not significant.
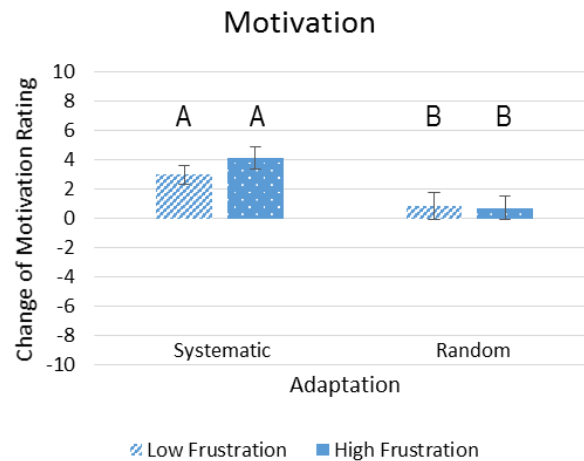


**Figure 24. Mean and standard error of motivation rating change (n=31).**

In order to investigate what happens to the other non-targeted learning factors, when targeting motivation, two 2x4 ANOVAs (one for systematic adaptation and one for random adaptation) were conducted with frustration and learning factor as variables. Tukey's post-hoc HSD analysis was conducted on the systematic and random adaptation conditions separately. In the systematic adaptation, the effect of learning factor was significant ($F(3,24)=31.2$, $p<.0001$, d=0.66). Frustration and the interaction were not significant. Figure 25 indicates significant pairwise differences for the systematic adaption. In the systematic condition, the change in motivation (the targeted variable) was significantly larger than the other learning factors. In the

random adaptation, no effect was significant. Figure 25 indicates no significant pairwise differences for the random adaption.



(a)                           (b)
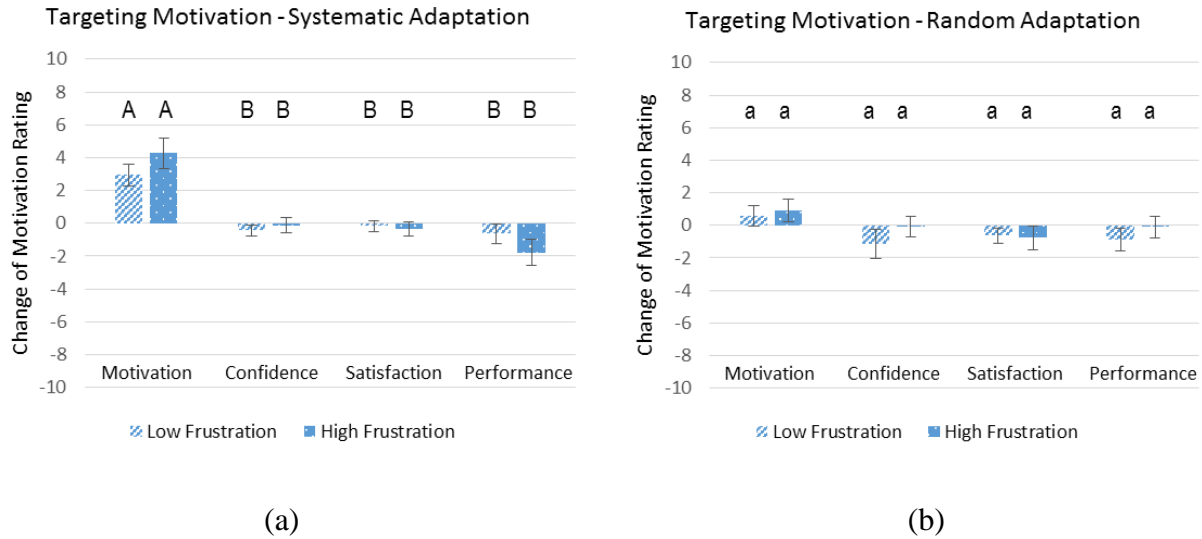
**Figure 25. Mean and standard error of motivation (a) with systematic (low frustration n=36, high frustration n=24) and (b) random adaptation (low frustration n=17, high frustration n=24).**

*Targeting Confidence.* In order to compare the effect of adaptation and frustration, a 2x2 ANOVA was conducted. Systematic adaptation resulted in a significantly ($F(1,15)=8.96$, $p=.0089$, d=0.35) larger gain in confidence than the random adaptation (see Figure 26). However, the effect of frustration, and the interaction between adaptation and frustration were not significant.

**Figure 26. Mean and standard error of confidence rating change (n=31).**

In order to investigate what happens to the other non-targeted learning factors, when targeting confidence, two 2x4 ANOVAs (one for systematic adaptation and one for random adaptation) were conducted with frustration and learning factor as variables. Tukey's post-hoc HSD analysis was conducted on the systematic and random adaptation conditions separately. In the systematic adaptation, the effect of learning factor was significant ($F(3,30)=18.7$, $p<.0001$, $d=0.41$). Frustration and the interaction were not significant. Figure 27a indicates significant pairwise differences for the systematic adaption. In the systematic condition, the change in confidence (the targeted variable) was significantly larger than the other learning factors. In the random adaptation, no effect was significant. Figure 27b indicates no significant pairwise differences for the random adaption.

(a)                                                                 (b)
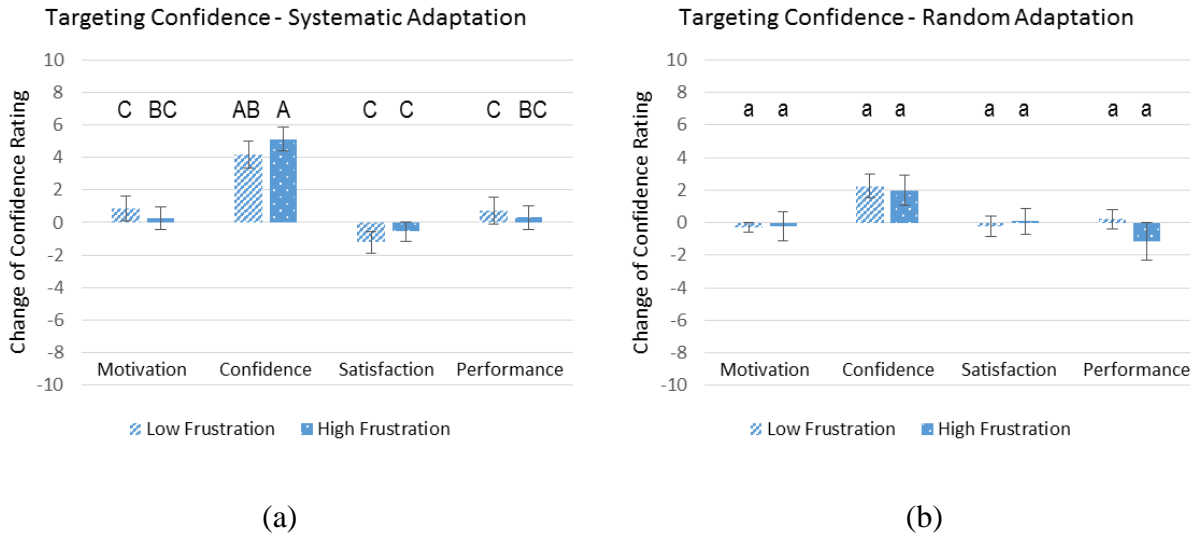
**Figure 27. Mean and standard error of confidence (a) with systematic (low frustration n=20, high frustration n=24) and (b) random adaptation (low frustration n=24, high frustration n=20).**

*Targeting Satisfaction.* In order to compare the effect of adaptation and frustration, a 2x2 ANOVA was conducted. Systematic adaptation resulted in a significantly ($F(1,17)=5.71$, *p=.0286*, d=0.16) larger gain in satisfaction than the random adaptation (see Figure 28). However, the effect of frustration, and the interaction between adaptation and frustration were not significant.
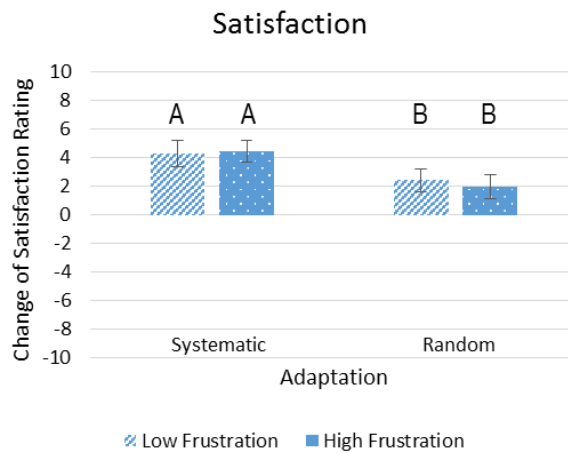


**Figure 28. Mean and standard error of satisfaction rating change (n=31).**

In order to investigate what happens to the other non-targeted learning factors, when targeting satisfaction, two 2x4 ANOVAs (one for systematic adaptation and one for random adaptation) were conducted with frustration and learning factor as variables. Tukey's post-hoc HSD analysis was conducted on the systematic and random adaptation conditions separately. In the systematic adaptation, the effect of learning factor was significant ($F(3,34)=17.1$, $p<.0001$, d=0.67). Frustration and the interaction were not significant. Figure 29a indicates significant pairwise differences for the systematic adaption. In the systematic condition, the change in satisfaction (the targeted variable) was significantly larger than the other learning factors. In the random adaptation, no effect was significant. Figure 29b indicates no significant pairwise differences for the random adaption.



(a)                                                                (b)
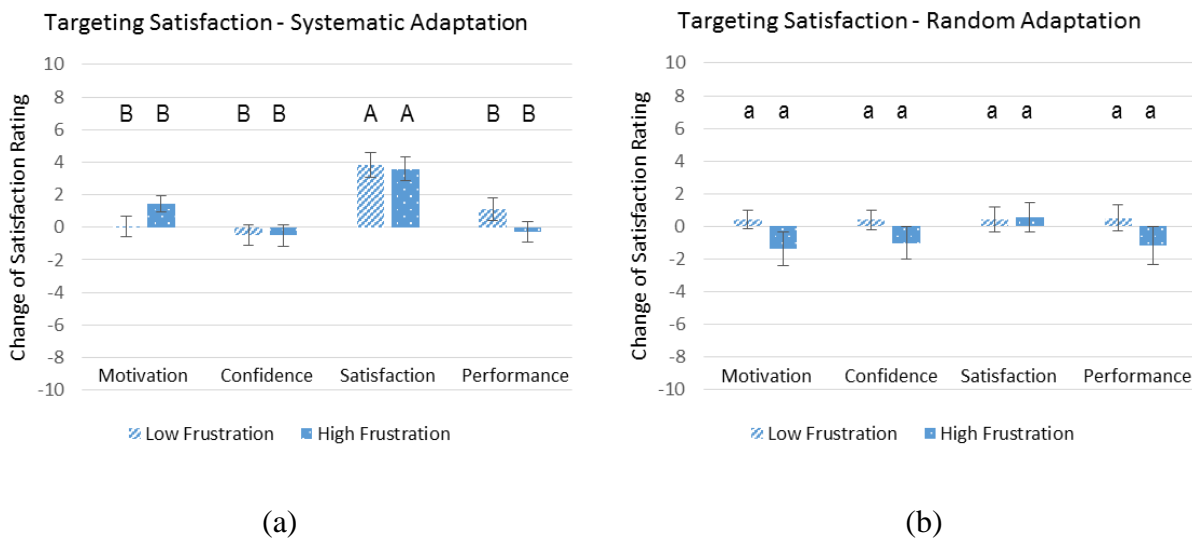
**Figure 29. Mean and standard error of satisfaction (a) with systematic (low frustration n=31, high frustration n=26) and (b) random adaptation (low frustration n=32, high frustration n=17).**

*Targeting Performance.* In order to compare the effect of adaptation and frustration, a 2x2 ANOVA was conducted. Systematic adaptation resulted in a significantly ($F(1,14)=10.3$, $p=.0061$, d=0.49) larger gain in performance than the random adaptation (see Figure 30).

However, the effect of frustration, and the interaction between adaptation and frustration were not significant.
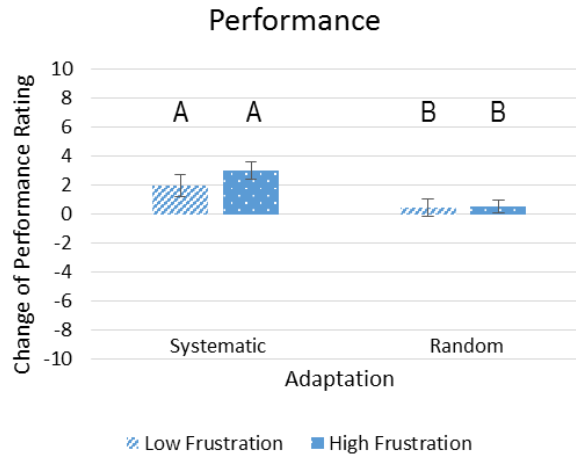


**Figure 30. Mean and standard error of performance rating change (n=31).**

In order to investigate what happens to the other non-targeted learning factors, when targeting performance, two 2x4 ANOVAs (one for systematic adaptation and one for random adaptation) were conducted with frustration and learning factor as variables. Tukey's post-hoc HSD analysis was conducted on the systematic and random adaptation conditions separately. In the systematic adaptation, the effect of learning factor was significant (F(3,44)=8.22, *p<.0002*, d=0.71). Frustration and the interaction were not significant. Figure 31a indicates significant pairwise differences for the systematic adaption. In the systematic condition, the change in performance (the targeted variable) was significantly larger than the 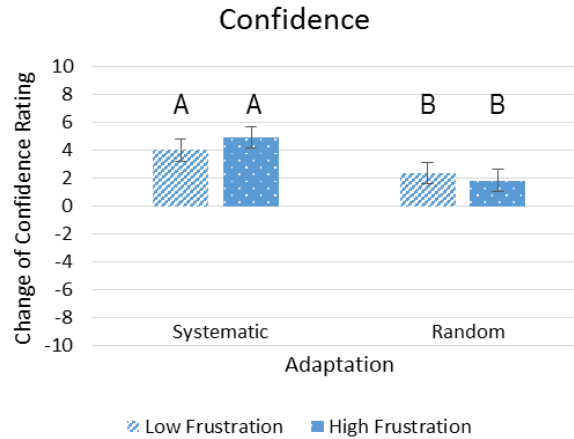other learning factors. In the random adaptation, no effect was significant. Figure 31b indicates no significant pairwise differences for the random adaption.
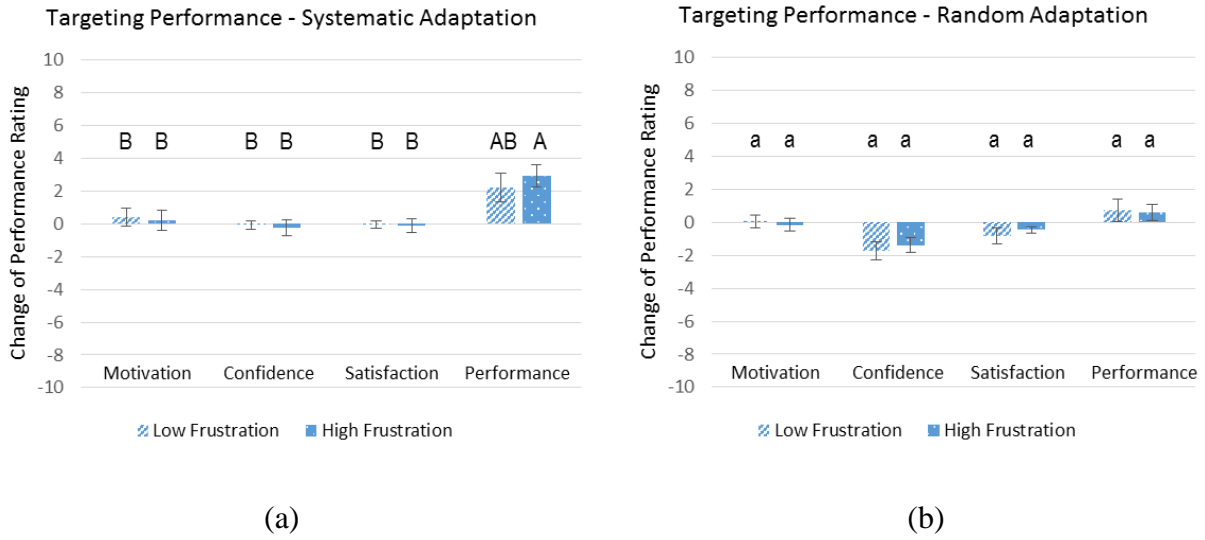
**Figure 31. Mean and standard error of performance (a) with systematic (low n=17, high n=30) and (b) random adaptation (low n=31, high n=43).**

## Feedback Appropriateness

The effect of adaptation and frustration individually were not significant. However, the interaction between adaptation and frustration was significant ($F(1,30)=13.9$, *p=.0008*, d=0.56). Figure 32 indicates significant pairwise differences between groups when they do not share a letter.
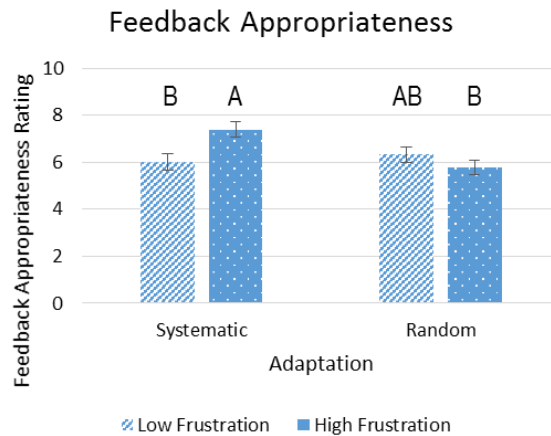


**Figure 32. Mean and standard error of feedback appropriateness (n=31).**

## Cognitive Workload

The effect of frustration was significant (F(1,30)=15.2, *p=.0005*, d=0.58) on cognitive workload. However, adaptation and the interaction between adaptation and frustration were not significant. Figure 33 indicates significant pairwise differences between groups when they do not share a letter.
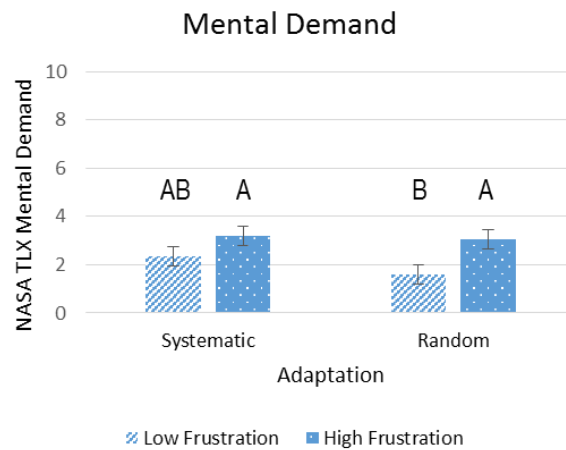


**Figure 33. Mean and standard error of TLX mental demand (n=31).**

## Stress

The effect of adaptation was significant (F(1,30)=17.5, *p=.0002*, d=0.61) on stress. The effect of frustration was also significant (F(1,30)=35.1, *p<.0001*, d=0.73). However, the interaction between adaptation and frustration were not significant. Figure 34 indicates significant pairwise differences between groups when they do not share a letter.
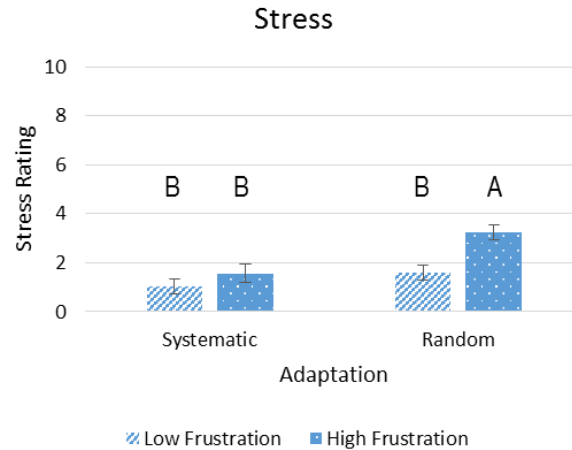
**Figure 34. Mean and standard error of stress (n=31).**

Discussion and Conclusions

The results of this study showed that an adaptive tutoring system prototype with systematic adaptation significantly influence motivation, confidence, satisfaction, and performance. The first hypothesis stated: "Systematic adaptation of etiquette strategies increase motivation, confidence, satisfaction, and performance more than random etiquette strategies." This hypothesis was supported. The level of motivation, confidence, and satisfaction was increased when systematic adaptation targeted to improve the participants' motivation, confidence, and satisfaction. Scores from the mathematics problems were higher when systematic adaptation targeted to increase the level of performance in both low and high frustration condition. On the other hand, feedback with random adaptation did not influence on motivation, confidence, satisfaction, and performance. These results showed that systematic adaptation is effective more than random etiquette strategies.

The second hypothesis stated: "Systematic adaptation of etiquette strategies mitigate user frustration more than random etiquette strategies." This hypothesis was fully supported. When

systematic adaptation was targeted to increase the level of motivation, confidence, satisfaction, and performance, all of them significantly improved in both high and low frustration condition. This demonstrates that systematic adaptation was always effective no matter what level of frustration the participants have. In addition, it shows that systematic adaptation was able to achieve same level of learning factor improvement no matter what level of frustration, thereby eliminating the effect of frustration as a negative factor in the learning process.

Besides using these results to check the hypotheses, this study also investigated the potential adaptation tradeoffs. The systematic adaptation system was able to improve the targeted leaning factor in each case (motivation, confidence, satisfaction, and performance). Furthermore, the targeted factor improvement was significantly larger than any of the non-targeted factor. Finally, the changes in the non-targeted factors were small, although sometimes they were negative. This is in contrast to the random adaptation system, where changes in all factors (targeted and non-targeted) were small and not significantly differ from each other.

Feedback with systematic adaptation when the participants were solving problems was rated more appropriate rather than random adaptation in high frustration condition. However, feedback appropriateness was rated as same between systematic and random adaptation in low frustration condition. This result shows that feedback can be more appropriate when the participants feel frustrated. Feedback with systematic adaptation also decrease the level of stress when the participants are highly frustrated. These results demonstrated that proper interaction style can mitigate the influences of frustration.

Results showed that providing the best interaction style of feedback depending on users' motivation, confidence, satisfaction, and performance can increase their learning processes. However, it does not mean that one strategy was always better across all four learning factors the

adaptation tradeoffs demonstrated this situation. It indicates that different strategies impact the learning factors in different ways. Since motivation, confidence, satisfaction, and performance are directly connected to the students' learning goals, providing appropriate feedback to support these is crucial to enhance effective learning (Keller, 1987). These results can be applicable for not only a human tutor but also a computer tutor.

In human-computer tutoring, most of the real-time adaptation is triggered by poor performance and results in a change to the task difficulty. However, a good human tutor will be aware of the emotional state of the learner and adapt their interaction style to support aspects of the student's learning that underlie performance such as a student's motivation, confidence, or satisfaction. Future work will fully develop an adaptive tutoring system depending on the problem-solving and emotional state of the students.

# CHAPTER VI: CONCLUSION

## Summary

Understanding the impact of feedback on user emotion and learning is important to human-computer collaboration applications. In Chapter II, a study on user emotion showed that time delay significantly influences physiological arousal, emotional states (frustration, anger), cognitive workload, and task performance. These results were true for both low and high task difficulties. Participants' reaction times to interruptions were slower, and they subjectively rated their mental demand higher in the time delay condition. In the post-experiment questionnaires, participants reported that they felt higher frustration and anger with the time delay than the no-delay condition. Participants experienced the lowest workload when they navigated without any delayed feedback. These results answer the first research question by showing how feedback influences on user emotions, cognitive workload, task performance, and physiological response.

Chapter III presented and discussed a study on interaction styles of feedback. The results of this study show that different feedback interaction styles impact different aspects of the learning process. For example, the participants performed better by receiving feedback based on bald and positive politeness under low frustration while they performed better with negative politeness feedback under high frustration. Their satisfaction with performance showed a similar pattern: participants were more satisfied when they received positive politeness feedback under low frustration, but negative politeness feedback under high frustration. These results demonstrated that different etiquette strategies were helpful to improve the participants' performances when they were highly frustrated. It provides the evidence that picking the proper interaction style can mitigate the influences of frustration. Likewise, the participants' ratings of motivation, satisfaction, and confidence showed a similar tendency. Thus a different etiquette

strategy may be appropriate depending on what factor is targeted for improvement; for instance, the etiquette strategy used to improve motivation in a high frustration situation is different than the strategy used to improve satisfaction in a low frustration period. These results answer the second research question by demonstrating that changing the interaction style of feedback based on etiquette strategies mitigate user frustration.

In Chapter IV, an adaptive tutoring system prototype was developed to investigate the effectiveness of adapting interaction styles of an intelligent tutoring system based on etiquette strategies results of Study 2. The results of this study showed that etiquette strategies could be effectively used in a systematic way to target improvement in areas of motivation, confidence, satisfaction, and performance in situations of low and high frustration. These results answer the third research question by showing that an adaptive system based on etiquette strategies mitigate user frustration and improve motivation, confidence, satisfaction, and performance.

Contribution

The results of these three studies provide contributions to the design of effective systems in HCI. First, this research investigated a new approach for designing feedback. This approach leveraged the concept of etiquette in linguistics from human-human interaction to HCI. The methodology of this study focuses on the interaction style between human and computer. This work demonstrated that different interaction styles, based on etiquette strategies, have differential effects on frustration, motivation, confidence, satisfaction, and performance.

Second, this study furthers the development of e human-like systems in HCI. Humans are finely attuned to the each other's emotions when they interact, and often the style by which they interact with another person is based on their assessment of the other person's emotions. In this

work, we developed a systematic approach towards enabling a computer system to modify its interaction style based on an assessment of the human's affective state. This study implemented an affect-aware adaptive system that is able to adjust its interaction styles.

Third, this paper developed a method to lessen the effects of user frustration. Previous work showed that frustration is a common problem in HCI and it decreases overall productivity. In this paper, the causes and impacts of user frustration were demonstrated as a result of delayed feedback. Furthermore, methods to mitigate user frustration were established by the application of etiquette strategies to positively affect human emotions.

Fourth, the results of this study demonstrated the ability to adapt interaction styles in HCI. Previous adaptive automation research has focused on primarily function allocation and to a lesser extent information content and task scheduling. However, the fourth approach, changing the interaction styles, is the least explored because of the interplay of human factors considerations. For instance, while changing the interaction style is a typical human trait when faced with certain situations, in HCI changing the way information is delivered to users may be a violation of the human factors principle of consistency (Feigh, Dorneich, & Hayes, 2012). This work focused exclusively on this fourth category and demonstrated that adapting interaction styles of a system can not only mitigate user frustration but also improve the learning process. More specifically, the methodology of adapting the interaction styles focused on changing the level of imposition by using etiquette strategies between human users and computer systems, mirroring human-human interaction. The results of this approach show that the level of imposition can also be changed in HCI to beneficial effect.

Fifth, the methods of this study designed and developed rules for dynamic feedback. To establish the rules, Study 2 demonstrated that different etiquette strategies have differential

effects. Study 3 enabled the system to target different learning states (i.e., motivation, confidence, satisfaction, performance) and improve them by using etiquette strategies.

Finally, this work developed and evaluated an affect-aware adaptive tutoring system. It mitigated the effects of user frustration and provided dynamic, systemic, targeted, improvement of users' motivation, confidence, satisfaction, and performance in the learning process. The system targeted not only the performance but also the supporting elements in the learning and problem-solving process, which has not typically been the emphasis of computer tutoring systems. Previous adaptive tutoring systems focused on adapting task difficulty and content because they only focused on overall performance in learning and HCI. This paper demonstrated a promising approach that utilized etiquette strategies in HCI to develop an affect-aware adaptive tutoring system.

Future Work

HCI researchers are still investigating the implementation and effects of the affect-aware adaptive systems. As such, many research questions remain open, including examination of the following topics: triggering of adaptations, diverse system domains, the scope of population, level of system fidelity, and various applications.

This work focused on testing the effectiveness of the adaptions and assumed perfect triggering. However, triggers can be important factors as well. For example, identifying the right way to detect the participants' errors or struggles during the six steps in problem-solving process needs to be verified. With regard to detection realm other issues need to be explored, such as: accuracy, individual differences, and learning styles. For instance, if a state will be automatically detected (e.g. frustration) through sensors and machine learning algorithms, what does the level

of accuracy have to be in order to adequately trigger adaptations. Study 1 used EDA to measure physiological arousal and FaceReader to detect user emotions. Although these tools are actively used to assess user states, the verification of detection accuracy needs to be clearly investigated. Comparison between subjective (e.g., questionnaire) and objective measurement (e.g., EDA) could be a way to examine the level of detection accuracy. Furthermore, an investigation that systemically explores different levels of accuracy in triggering, and assess the impact on the effectiveness of the adaptations in the closed loop adaptive system, will help determine triggering accuracy thresholds.

Individual differences of the participants need to be considered. Tasks from Study 2 and Study 3 required an ability to solve certain level of mathematics problems. However, each individual has different level of such problem-solving ability. Although Study 2 and Study 3 recruited only people from in STEM fields to mitigate this limitation, considering individual differences still needs to be examined to generalize the findings of this work. Furthermore, the population used in Study 2 and Study 3 were students, who are used to taking tests and solving problems. Adult learners, or non-student populations, may not be as habituated to the testing paradigm, and thus may need other styles of tutoring to support for effective learning.

Currently, many affect-aware systems are focused on the learning domain. However affect aware systems may have a wider application infields that include a human user as a part of the system (e.g., aviation, military). Testing the affect-aware adaptive systems in diverse domains may provide promising approaches to obtain better human-machine communications. Likewise, evaluations of a wider population scope of the participants (e.g., age, expertise, circumstances, gender, background, culture) can expand the applicability and impact of affect-aware adaptive systems. In this study, the task was only mathematics problems and targeted

current University students. A future area of research would be to expand this adaptive interaction style approach to non-STEM fields. In a similar manner, if the task is not about mathematics, then different approaches to develop the feedback and providing customized training session need to be investigated.

This work recruited only students as participants in school. However, people who are not in school (e.g., adult learner, employee in company) need to be considered. Depending on the contents they learn, they might not be accustomed to taking tests and quizzes during the learning phase. In this case, the feedback may need to be different and customized based on the circumstances.

Different genders of the participants may need to be considered as well. There are different phenomenon between male and female students in STEM fields. For example, it has been shown that the level of confidence, and interest on STEM contents of female students were undermined (Booth & Gerard, 2011; Shapiro & Williams, 2012; Heaverlo, Cooper, & Lannan, 2013; Sobel, Gilmartin, & Sankar, 2016) although the level of performance is same as male students in some male-dominated fields such as STEM (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). In this situation, the same approach of feedback for both male and female may not be appropriate because lack of confidence leads to lower motivation. For solving this problem, different types of feedback need to be developed by considering the gender differences of the students in STEM fields. This work may lay the foundation for addressing a wider range of learning factors (e.g. confidence) that may lead to ways to mitigate gender differences in the design of adaptive tutoring systems.

Finally, a higher level of fidelity for the affect-aware adaptive systems need to be implemented and tested to produce more usable systems for users. Such future studies will allow

the HCI community to approach ways to create more useful, effective, and applicable affect-

aware adaptive systems to support people who use computers.

BIBLIOGRAPHY

Allison, R. S., Harris, L. R., Jenkin, M., Jasiobedzka, U., & Zacher, J. E. (2001, March). Tolerance of temporal delay in virtual environments. *In Virtual Reality*, 2001. Proceedings. IEEE (pp. 247-254). IEEE.

Allison, R. S., Zacher, J. E., Wang, D., & Shu, J. (2004, June). Effects of network delay on a collaborative motor task with telehaptic and televisual feedback. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry* (pp. 375-381). ACM.

Amsel, A. (1992). Frustration theory: An analysis of dispositional learning and memory (No. 11). *Cambridge University Press*.

Arcara, P., & Melchiorri, C. (2002). Control schemes for teleoperation with time delay: A comparative study. *Robotics and Autonomous Systems*, 38(1), 49-64.

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B.P. (2007). Repairing disengagement with non-invasive interventions. *AIED*.195-202.

Asada, M. (2014). Towards artificial empathy. *International Journal of Social Robotics*, 7(1), 19-33.

Ash, A., Palmisano, S., Govan, D. G., & Kim, J. (2011). Display lag and gain effects on vection experienced by active observers. *Aviation, space, and environmental medicine*, 82(8), 763-769.

Azevedo, R., Witherspoon, A., Chauncey, A., Burkett, C., & Fike, A. (2009, October). MetaTutor: A MetaCognitive tool for enhancing self-regulated learning. In *2009 AAAI Fall Symposium Series*.

Azevedo, R., Witherspoon, A. M., Graesser, A. C., McNamara, D. S., Chauncey, A., Siler, E., Cai, Z., Rus, V., & Lintean, M. C. (2009, July). MetaTutor: Analyzing Self-Regulated Learning in a Tutoring System for Biology. In *AIED* (pp. 635-637).

Back, A.L., Arnold, R.M., Baile, W.F., Fryer-Edwards, K.A., Alexander, S.C., Barley, G.E., Gooley, T.A., & Tulsky, J.A. (2007). Efficacy of communication skills training for giving

bad news and discussing transitions to palliative care. *Archives of internal medicine*, 167(5), 453-460.

Baker, R.S., Corbett, A.T., & Koedinger, K.R. (2004). Detecting student misuse of intelligent tutoring systems. Intelligent tutoring systems (pp. 531-540). Springer Berlin Heidelberg.

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality*, 40(1), 35-55.

Barrowcliff, A. L., Gray, N. S., Freeman, T. C., & MacCulloch, M. J. (2004). Eye-movements reduce the vividness, emotional valence and electrodermal arousal associated with negative autobiographical memories. *Journal of Forensic Psychiatry & Psychology*, 15(2), 325-345.

Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and cognition*, 55(1), 30-40.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, 10(3), 295-307.

Behnke, S., Egorova, A., Gloye, A., Rojas, R., & Simon, M. (2004). Predicting away robot control latency. In *RoboCup 2003: Robot soccer world cup VII* (pp. 712-719). Springer Berlin Heidelberg.

Bellon, J. J., Bellon, E. C., & Blank, M. A. (1992). Teaching from a research knowledge base: A development and renewal process. *Merrill*.

Benţa, K. I., Van Kuilenburg, H., Eligio, U. X., Den Uyl, M., Cremene, M., Hoszu, A., & Creţ, O. (2009). Evaluation of a System for RealTime Valence Assessment of Spontaneous Facial Expressions. In *Distributed Environments Adaptability, Semantics and Security Issues International Romanian-French Workshop*, Cluj-Napoca, Romania (pp. 17-18).

Bessiere, K., Newhagen, J. E., Robinson, J. P., & Shneiderman, B. (2006). A model for computer frustration: The role of instrumental and dispositional factors on incident, session, and post-session frustration and mood. *Computers in human behavior*, 22(6), 941-961.

Bidwell, J., Holloway, A., & Davidoff, S. (2014, April). Measuring operator anticipatory inputs in response to time-delay for teleoperated human-robot interfaces. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1467-1470). ACM.

Booth, M. Z., & Gerard, J. M. (2011). Self-esteem and academic achievement: a comparative study of adolescent students in England and the United States. *Compare: A Journal of Comparative and International Education*, 41(5), 629-648.

Boucsein, W. (2012). Electrodermal activity. *Springer Science & Business Media*.

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. *Cognitive neuroscience of emotion*, 25, 49-59.

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. *Psychophysiology*, 49, 1017-1034.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). How people learn: Brain, mind, experience, and school. Washington, DC: *National Academy Press*.

Broderick, Z. (2011). Increasing parent engagement in student learning using an Intelligent Tutoring System with Automated Messages (Doctoral dissertation, *Worcester Polytechnic Institute*).

Brown, J. A. (1954). The social psychology of industry.

Calvo, R. & D'Mello, S. (2012). Frontiers of affect-aware learning technologies. *Intelligent Systems, IEEE*, 27(6), 86-89.

Card, S. K., Robertson, G. G., & Mackinlay, J. D. (1991). The information visualizer, an information workspace. In *Proceedings of the SIGCHI Conference on Human factors in computing systems* (pp. 181-186). ACM.

Carlson, N. R. (2013). Physiology of behavior. *Pearson*.

Casals, A. (1998). Robots in surgery. In *Autonomous Robotic Systems* (pp. 222-234). Springer London.

Chen, J.S., Gross, K., & Amsel, A. (1981). Ontogeny of successive negative contrast and its dissociation from other paradoxical reward effects in preweanling rats. *Journal of Comparative and Physiological Psychology*, 95(1), 146.

Chen, T. L., King, C. H. A., Thomaz, A. L., & Kemp, C. C. (2014). An investigation of responses to robot-initiated touch in a nursing context. *International Journal of Social Robotics*, 6(1), 141-161.

Christie, M. J. (1981). Electrodermal activity in the 1980s: a review. *Journal of the Royal Society of Medicine*, 74(8), 616.

Conesa, J. (1995). Electrodermal palmar asymmetry and nostril dominance. *Perceptual and motor skills*, 80(1), 211-216.

Conklin, J. E. (1957). Effect of control lag on performance in a tracking task. *Journal of experimental psychology*, 53(4), 261.

Cootes, T. F., & Taylor, C. J. (2004). Statistical models of appearance for computer vision. Technical report, University of Manchester, Wolfson Image Analysis Unit, *Imaging Science and Biomedical Engineering*.

Corbett, A. T., & Anderson, J. R. (2008). Student Modeling and Mastery Learning in a Computer-Based Programming Tutor. *Department of Psychology*, 18.

Corde Lane, J., Carignan, C. R., Sullivan, B. R., Akin, D. L., Hunt, T., & Cohen, R. (2002). Effects of time delay on telerobotic control of neutral buoyancy vehicles. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on* (Vol. 3, pp. 2874-2879). IEEE.

D'Arcey, J. T. (2013). Assessing the validity of FaceReader using facial EMG (Doctoral dissertation, *California State University*, Chico).

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). 7 The Electrodermal System. *Handbook of psychophysiology*, 159.

Dempsey, J. V. (1993). Interactive instruction and feedback. *Educational Technology*.

De Silva, L. C., Miyasato, T., & Nakatsu, R. (1997). Facial emotion recognition using multi-modal information. In *Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on* (Vol. 1, pp. 397-401). IEEE.

De Silva, L. C., & Ng, P. C. (2000). Bimodal emotion recognition. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (pp. 332-335). IEEE.

Den Uyl, M.J.; Van Kuilenburg, H. (2008). The FaceReader: Online Facial Expression Recognition. *Proceedings of Measuring Behavior 2005*, Wageningen, The Netherlands, August 30 - September 2, 2008, pp. 589-590.

Dickman, I. R., & Gordon, S. (1985). One Miracle at a Time: How to Get Help for Your Disabled Child-from the Experiences of Other Parents. *Simon & Schuster*.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R. and Graesser, A., (2008). AutoTutor detects and responds to learners affective and cognitive states. In *Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems* (pp. 306-308).

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). Frustration and aggression. *Yale University Press*, New Haven

Dorneich, M. C., Ververs, P. M., Mathan, S., Whitlow, S., & Hayes, C. C. (2012). Considering etiquette in the design of an adaptive system. *Journal of Cognitive Engineering and Decision Making*, 6(2), 243-265.

Draper, J. V. (1993, December). Human factors in telemanipulation: perspectives from the Oak Ridge National Laboratory experience. In *Optical Tools for Manufacturing and Advanced Automation* (pp. 162-174). International Society for Optics and Photonics.

Drascic, D., Milgram, P., & Grodski, J. (1989, November). Learning effects in telemanipulation with monoscopic versus stereoscopic remote viewing. In *Systems, Man and Cybernetics, 1989. Conference Proceedings, IEEE International Conference on* (pp. 1244-1249). IEEE.

Duncan, S., & Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition and emotion*, 21(6), 1184-1211.

Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research* Digest, 8, 151-158.

Ellis, S. R., Mania, K., Adelstein, B. D., & Hill, M. I. (2004, September). Generalizeability of latency detection in a variety of virtual environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 23, pp. 2632-2636). SAGE Publications.

Fabrlzio, M. D., Lee, B. R., Chan, D. Y., Stoianovici, D., Jarrett, T. W., Yang, C., & Kavoussi, L. R. (2000). Effect of time delay on surgical performance during telesurgical manipulation. *Journal of endourology*, 14(2), 133-138.

Feigh, K.M., Dorneich, M.C., & Hayes, C.C. (2012). "Towards a Characterization of Adaptive Systems: a Framework for Researchers and System Designers", *Journal of Human Factors and Ergonomics*. Vol. 54, No. 6, pp. 1008-1024.

Ferdig, R. E., & Mishra, P. (2004). Emotional Responses to Computers: Experiences in Unfairness, Anger, and Spite1. *Journal of Educational Multimedia and Hypermedia*, 13(2), 143.

Fisher, C. D., & Noble, C. S. (2004). A within-person examination of correlates of performance and emotions while working. *Human Performance*, 17(2), 145-168.

Fogg, B. J. (1998). Persuasive computers: perspectives and research directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 225-232). ACM Press/Addison-Wesley Publishing Co.

Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002 (December).

Foley, W. A. (1997). Anthropological linguistics. *Blackwell Publishing Ltd*.

Frijda, N. H. (1986). The emotions. *Cambridge University Press*.

Gamalel-Din, S. (2002, June). The Smart Tutor: Student-Centered Case-Based Adaptive Intelligent e-Tutoring. In *the Proceedings of the 1st International Conference on Informatics and Systems*, Cairo (Vol. 17, p. 20).

Geister, S., Konradt, U., & Hertel, G. (2006). Effects of process feedback on motivation, satisfaction, and performance in virtual teams. *Small group research*, 37(5), 459-489.

Gelman, A. (2013). Commentary: P values and statistical practice. *Epidemiology*, 24(1), 69-72.

Gilbert, S. B., Blessing, S. B., & Guo, E. (2015). Authoring Effective Embedded Tutors: An Overview of the Extensible Problem Specific Tutor (xPST) System. *International Journal of Artificial Intelligence in Education*, 1-27.

Goldberg, B. S., Holden, H. K., Brawner, K. W., & Sottilare, R. A. (2011). Enhancing performance through pedagogy and feedback: Domain considerations for ITSs. In *Interservice/Industry Training, Simulation, and Education Conference*, I/ITSEC.

Goffman, E. (1967). On face-work. *Interaction ritual*, 5-45.

Gordon, T. (2008). Parent effectiveness training: The proven program for raising responsible children. *Harmony*.

Gordon, T., & Bruch, N. (1974). Teacher effectiveness training. *New York: PH Wyden*.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. Education, *IEEE Transactions on*, 48(4), 612-618.

Granić, A., Stankov, S., & Glavinić, V. (2000, June). User interface aspects of an intelligent tutoring system. In Information Technology Interfaces, 2000. ITI 2000. *Proceedings of the 22nd International Conference on* (pp. 157-164). IEEE.

Hansen, S., & Eddy, E. (2007). Engagement and frustration in programming projects. *ACM SIGCSE Bulletin*. 39(1), 271-275. ACM.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, 48, 615-625.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.

Hattie, J. (2008). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. *Routledge*.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.

Hayes, C. C. & Miller, C. A. (2010). Should computers be polite? In C. C. Hayes & C. A. Miller (Eds.), Human-computer etiquette: Cultural expectations and the design implications they place on computers and technology (pp. 1–12). *Boca Raton, FL: Taylor & Francis*

Hazlett, R. (2003). Measurement of user frustration: a biologic approach. In *CHI'03 extended abstracts on Human factors in computing systems* (pp. 734-735). ACM.

Heaverlo, C. A., Cooper, R., & Lannan, F. S. (2013). STEM development: Predictors for 6th-12th grade girls' interest and confidence in science and math. *Journal of Women and Minorities in Science and Engineering*, 19(2)

Held, R., & Durlach, N. (1991). Telepresence, time delay and adaptation. *Pictorial communication in virtual and real environments*, 232-246.

Held, R., Efstathiou, A., & Greene, M. (1966). Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology*, 72(6), 887.

Hokanson, J.E., & Burgess, M. (1964). Effects of physiological arousal level, frustration, and task complexity on performance. *The Journal of Abnormal and Social Psychology*, 68(6), 698.

Hone, K. (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers*, 18(2), 227-245.

Hoxmeier, J. A., & DiCesare, C. (2000). System response time and user satisfaction: An experimental study of browser-based applications. *AMCIS 2000 Proceedings*, 347.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, 321(5888), 494-495.

Jeon, M., Walker, B.N., & Yim, J.B. (2014). Effects of specific emotions on subjective judgment, driving performance, and perceived workload. *Transportation research part F: traffic psychology and behavior*, 24, 197-209.

Jia, Y., Xi, N., Liu, S., Wang, Y., Li, X., & Bi, S. (2014). Quality of teleoperator adaptive control for telerobotic operations. *The International Journal of Robotics Research*, 33(14), 1765-1781.

Johnson, W.L., Friedland, L., Schrider, P., Valente, A., & Sheridan, S. (2011). The Virtual Cultural Awareness Trainer (VCAT): Joint Knowledge Online's (JKO's) solution to the individual operational culture and language training gap. *Proceedings of ITEC*.

Kaliouby, R., Picard, R., & Baron-Cohen, S. (2006). Affective computing and autism. *Annals of the New York Academy of Sciences*, 1093(1), 228-248.

Kasper, G. (2005). Linguistic etiquette. *Intercultural discourse and communication: The essential readings*, 58-67.

Kędzierski, J., Muszyński, R., Zoll, C., Oleksy, A., & Frontkiewicz, M. (2013). EMYS— emotive head of a social robot. *International Journal of Social Robotics*, 5(2), 237-249.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of instructional development*, 10(3), 2-10.

Keller, J. M. (2009). Motivational design for learning and performance: The ARCS model approach. *Springer Science & Business Media*.

Keller, J. M., & Kopp, T. W. (1987). An application of the ARCS Model of Motivational Design.

Keller, J. M., & Suzuki, K. (1988). Use of the ARCS Motivation Model in courseware design.

Kiesler, S., Zubrow, D., Moses, A. M., & Geller, V. (1985). Affect in computer-mediated communication: An experiment in synchronous terminal-to-terminal discussion. *Human-Computer Interaction*, 1(1), 77-104.

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing*, 42(3), 419-427.

Klein, J., Moon, Y., & Picard, R. W. (1999). This computer responds to user frustration. In *CHI'99 extended abstracts on Human factors in computing systems* (pp. 242-243). ACM.

Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with computers*, 14(2), 119-140.

Klein, T., Rosalind, C., & Picard, W. (1999). Computer response to user frustration.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city.

Kort, B., & Reilly, R. (2002, January). An affective module for an intelligent tutoring system. In *Intelligent Tutoring Systems* (pp. 955-962). Springer Berlin Heidelberg.

Kort, B., Reilly, R., & Picard, R. W. (2001, August). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In *icalt* (p. 0043). IEEE.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance*, 279-328.

Kulic D, Croft E (2007) Pre-collision safety strategies for human robot interaction. *Auton Robots* 22(2):149–164

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3), 261-273.

Lang P (1995) The emotion probe: studies of motivation and attention. Am *Psychol* 50(5):372

Lazar, J., Jones, A., Hackley, M., & Shneiderman, B. (2006). Severity and impact of computer user frustration: A comparison of student and workplace users. *Interacting with Computers*, 18(2), 187-207.

Lazar, J., Jones, A., & Shneiderman, B. (2006). Workplace user frustration with computers: An exploratory investigation of the causes and severity. *Behaviour & Information Technology*, 25(03), 239-251.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Psychology*, 66.

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, 7(4), 227.

Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. Chae, S.-Kang, U.–Jeon, E.–Linacre, JM (eds.): Development of Computerized Middle School Achievement Tests, *MESA Research Memorandum*, (69).

Li, L., & Chen, J. H. (2006). Emotion recognition using physiological signals. In *Advances in Artificial Reality and Tele-Existence* (pp. 437-446). Springer Berlin Heidelberg.

Liu, C. L. & Yu, Y. (2011). The Research Summaries on The Pressure and The Frustration of Undergraduate Employment. *Journal of Jilin Normal University (Humanities & Social Science Edition)*, 1, 005.

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R.S. (2013). Sequences of Frustration and Confusion, and Learning. *EDM* (114-120).

Lindquist, K. A., & Barrett, L. F. (2008). Emotional complexity. *Handbook of emotions*, 513-530.

Loijens, L., Krips, O., Van Kuilenbug, H., Den Uyl, M., Ivan, P., Theuws, H., and Spink, A. (2012) FaceReader Reference Manual Version 5.0, *Noldus Information Technology b.v.*

London, M. (2003). Job feedback: Giving, seeking, and using feedback for performance improvement. *Psychology Press*.

Luck, J. P., McDermott, P. L., Allender, L., & Russell, D. C. (2006, March). An investigation of real world control of robotic assets under communication latency. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 202-209). ACM.

Lum, M. J., Rosen, J., Lendvay, T. S., Sinanan, M. N., & Hannaford, B. (2009, May). Effect of time delay on telesurgical performance. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on* (pp. 4246-4252). IEEE.

Maaoui, C., Pruski, A., & Abdat, F. (2010). Emotion recognition through physiological signals for human-machine communication. *Cutting Edge Robotics*, 317-332.

Magalie Ochs, M., & Frasson, C. (2004). Emotionally Intelligent Tutoring Systems (EITS). *Association for the Advancement of Artificial Intelligence*.

Malik, S. (2014). Effectiveness of Arcs Model of Motivational Design to Overcome Non Completion Rate of Students in Distance Education. *Turkish Online Journal of Distance Education*, 15(2).

Martin, G. N., Hogg, M. A., & Abrams, D. (2010). Social cognition and attitudes. *Journal of psychology, Harlow: Pearson education limited, page*, (646-677).

Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). Classroom instruction that works: Research-based strategies for increasing student achievement. *Ascd*.

Meehan, M., Razzaque, S., Whitton, M. C., & Brooks Jr, F. P. (2003, March). Effect of latency on presence in stressful virtual environments. In *Virtual Reality, 2003. Proceedings. IEEE* (pp. 141-148). IEEE.

McGill, I., & Brockbank, A. (2003). Action learning handbook. *Kogan Page*.

Meisner E, Isler V, Trinkle J (2008) Controller design for humanrobot interaction. *Auton Robots* 24:123–134

Mills, S. (2003). Gender and politeness (Vol. 17). *Cambridge University Press.*

Miller, C.A. & Dorneich, M.C. (2006), "From Associate Systems to Augmented Cognition:  25 Years of User Adaptation in High Criticality Systems", *Proceedings of the 2nd Augmented Cognition International)*, San Francisco, CA.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(1), 57-75.

Miller, C. A., Wu, P., & Funk, H. B. (2008). A computational approach to etiquette: Operationalizing Brown and Levinson's politeness model. Intelligent Systems, *IEEE*, 23(4), 28-35.

Miller, N. E. (1941). I. The frustration-aggression hypothesis. *Psychological Review*, 48(4), 337.

Miller, R. B. (1968). Response time in man-computer conversational transactions. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I* (pp. 267-277). ACM.

Mohammad, S., & Job, M.A. (2012). Confidence-Motivation–Satisfaction-Performance (CMSP) Analysis of Blended Learning System in the Arab Open University Bahrain. *International Journal of Information Technology and Business Management*, 3(1).

Morlan, G. K. (1949). A note on the frustration-aggression theories of Dollard and his associates. *Psychological review*, 56(1), 1.

Murray, T. (2003). An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art. In *Authoring tools for advanced technology learning environments* (pp. 491-544). Springer Netherlands.

Nasoz, F., Alvarez, K., Lisetti, C. L., & Finkelstein, N. (2004). Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology & Work*, 6(1), 4-14.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates?. *International Journal of Human-Computer Studies*, 45(6), 669-678.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities?. *International Journal of Human-Computer Studies*, 43(2), 223-239.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78). ACM.

Nielsen, J. (1994). Usability engineering. *Elsevier*.

Nkambou, R. (2010). Advances in intelligent tutoring systems (Vol. 308). *Springer Science & Business Media.*

Nomura, T., & Nakao, A. (2010). Comparison on identification of affective body motions by robots between elder people and university students: a case study in Japan. *International Journal of Social Robotics*, 2(2), 147-157.

Nonami, K., Shimoi, N., Huang, Q. J., Komizo, D., & Uchida, H. (2000). Development of teleoperated six-legged walking robot for mine detection and mapping of mine field. In Intelligent Robots and Systems, 2000.(IROS 2000). *Proceedings. 2000 IEEE/RSJ International Conference on* (Vol. 1, pp. 775-779). IEEE.

Norcini, J. (2010). The power of feedback. *Medical education*, 44(1), 16-17.

Octavia, J. R., Raymaekers, C., & Coninx, K. (2011). Adaptation in virtual environments: conceptual framework and user models. *Multimedia Tools and Applications*, 54(1), 121-142.

Ohba, K., Clary, G., Hiratuka, S., Tsukada, T., Kotoku, T., & Tanie, K. (1998, October). Gesture and facial expression on tele-robotics. In Intelligent Robots and Systems, *Proceedings of IEEE/RSJ International Conference on* (Vol. 3, pp. 1596-1601). IEEE.

Ong, J., & Ramachandran, S. (2003). Intelligent tutoring systems: Using ai to improve training performance and roi. *Networker Newsletter*, 19(6).

Owen-Hill, A., Suárez-Ruiz, F., Ferre, M., & Aracil, R. (2014, January). Effect of Video Quality and Buffering Delay on Telemanipulation Performance. In *ROBOT2013: First Iberian Robotics Conference* (pp. 555-568). Springer International Publishing.

Padayachee, I. (2002). Intelligent tutoring systems: Architecture and characteristics. *University of Natal, Durban, Information Systems & Technology, School of Accounting & Finance*.

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, 55(9), 1059-1072.

Parasuraman, R., & Miller, C.A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM*, 47(4), 51-55.

Patten, C. J., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: cognitive workload and attention resource allocation. *Accident analysis & prevention*, 36(3), 341-350.

Pearson, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and instruction*, 13(2), 161-188.

Pérez-Quiñones, M. A., & Sibert, J. L. (1996, April). A collaborative model of feedback in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 316-323). ACM.

Picard, R. W. (2006). Building an affective learning companion. In *Intelligent Tutoring Systems* (pp. 811-811).

Picard, R. W., & Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications. *Interacting with computers*, 14(2), 141-169.

Picard, R.W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M, Deb Roy, and Carol Strohecker. (2004). Affective learning—a manifesto. *BT Technology Journal*, 22(4), 253-269.

Ponto, K., Kohlmann, J., & Tredinnick, R. (2014). DSCVR: designing a commodity hybrid virtual reality system. *Virtual Reality*, 19(1), 57-70.

Powers, S. R., Rauh, C., Henning, R. A., Buck, R. W., & West, T. V. (2011). The effect of video feedback delay on frustration and emotion communication accuracy. *Computers in Human Behavior*, 27(5), 1651-1657.

Prewett, M. S., Johnson, R. C., Saboe, K. N., Elliott, L. R., & Coovert, M. D. (2010). Managing workload in human–robot interaction: A review of empirical studies. *Computers in Human Behavior*, 26(5), 840-856.

Qu, L., Wang, N., & Johnson, W.L. (2005). Using learner focus of attention to detect learner motivation factors. *User Modeling 2005* (70-73). Springer Berlin Heidelberg.

Radin, D. I. (2004). Electrodermal presentiments of future emotions. *Journal of Scientific Exploration*, 18(2), 253-273.

Renaud, K., & Cooper, R. (2000). Feedback in human-computer interaction-characteristics and recommendations. *South African Computer Journal*, 26, 105-114.

Ridao, P., Carreras, M., Hernandez, E., & Palomeras, N. (2007). Underwater telerobotics for collaborative research. In *Advances in Telerobotics* (pp. 347-359). Springer Berlin Heidelberg.

Riek, L. D., & Robinson, P. (2009). Affective-centered design for interactive robots. In *Proceedings of the AISB symposium on new frontiers in human-robot interaction.*

Robert, G., & Hockey, J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology*, 45(1), 73-93.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, 21(2), 267-280.

Rosenthal-von der Pütten AM, Krämer NC, Hoffmann L, Sobieraj S, Eimler SC (2013) An experimental study on emotional reactions towards a robot. *Int J Soc Robot* 5(1):17–34.

Rowe, D. W., Sibert, J., & Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 480-487). ACM Press/Addison-Wesley Publishing Co.

Salman, A. R. (2013). The Use of Intelligent Tutoring System for Developing Web-based Learning Communities. *IJCSI International Journal of Computer Science Issues*, 10(6).

Sands, W. A., Waters, B. K., & McBride, J. R. (1997). Computerized adaptive testing: From inquiry to operation. *American Psychological Association*.

Satava, R. M., & Simon, I. B. (1993). Teleoperation, telerobotics, and telepresence in surgery. *Endoscopic surgery and allied technologies*, 1(3), 151-153.

Scerbo, M. W.; Freeman, F. G. & Mikulka, P. J. (2003). "A brain-based system for adaptive automation". *Theoretical Issues in Ergonomic Science*, 4, 200-219.

Scheirer, J., Fernandez, R., Klein, J., & Picard, R. W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with computers*, 14(2), 93-118.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological review*, 61(2), 81.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, 14(4), 433-440.

Sears, R. R. (1937). Initiation of the repression sequence by experienced failure. *Journal of Experimental Psychology*, 20(6), 570.

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, 66(3-4), 175-183.

Sheik-Nainar, M. A., Kaber, D. B., & Chow, M. Y. (2005). Control gain adaptation in virtual reality mediated human–telerobot interaction. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 15(3), 259-274.

Shellnut, B., Knowltion, A., & Savage, T. (1999). Applying the ARCS model to the design and development of computer-based modules for manufacturing engineering courses. *Educational technology research and development*, 47(2), 100-110.

Sheridan, T. B. (1992). Telerobotics, automation, and human supervisory control. *MIT press*.

Sheridan, T. B., & Ferrell, W. R. (1963). Remote manipulative control with transmission delay. *IEEE Transactions on Human Factors in Electronics*, 4(1), 25-29.

Shneiderman, B. (2003). Designing the user interface. *Pearson Education India*.

Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1), 140-164.

Sobel, M., Gilmartin, J., & Sankar, P. (2016). Class Size and Confidence Levels Among Female STEM Students [Impact]. *IEEE Technology and Society Magazine*, 35(1), 23-26.

Solkoff, N., Todd, G. A., & Screven, C. G. (1964). Effects of frustration on perceptual-motor performance. *Child Development*, 569-575.

Sollenberger, D. J., & Singh, M. P. (2012). Koko: an architecture for affect-aware games. *Autonomous Agents and Multi-Agent Systems*, 24(2), 255-286.

Spector, P. E. (1975). Relationships of Organizational Frustration with Reported Behavioral Reactions of Employees. *Journal of Applied Psychology*, 60(5), 635-637.

Swangnetr M, Zhu B, Kaber D, Taylor K (2010) Meta-analysis of user age and service robot configuration effects on human-robot interaction in a healthcare application. In: *2010 AAAI fall symposium series.*

Sylwester, R. (1994). How emotions affect learning. *Educational Leadership*, 52(2), 60-65.

Szameitat, A. J., Rummel, J., Szameitat, D. P., & Sterr, A. (2009). Behavioral and emotional consequences of brief delays in human–computer interaction. *International Journal of Human-Computer Studies*, 67(7), 561-570.

Toda, M. (1980). Emotion and decision making. *Acta Psychologica*, 45(1), 133-155.

Van der Linden, W. J., & Glas, C. A. (2000). Computerized adaptive testing: Theory and practice. *Dordrecht: Kluwer Academic*.

VanLehn, K., Burleson, W., Girard, S., Chavez-Echeagaray, M. E., Gonzalez-Sanchez, J., Hidalgo-Pontet, Y., & Zhang, L. (2014, June). The affective meta-tutoring project: lessons learned. In *Intelligent Tutoring Systems* (pp. 84-93). Springer International Publishing.

Visser, J., & Keller, J. M. (1990). The clinical use of motivational messages: An inquiry into the validity of the ARCS model of motivational design. *Instructional science*, 19(6), 467-500.

Voeffray, S. (2011). Emotion-sensitive Human-Computer Interaction (HCI): State of the art-Seminar paper. *Emotion Recognition*. p1-4.

Wahlström, J., Hagberg, M., Johnson, P., Svensson, J., & Rempel, D. (2002). Influence of time pressure and verbal provocation on physiological and psychological reactions during work with a computer mouse. *European journal of applied physiology*, 87(3), 257-263.

Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B. F., Mislevy, R. J., & Steinberg, L. (2001). Computerized adaptive testing: A primer. *Qual Life Res*, 10(8), 733-734.

Wainer, H., & Mislevy, R. J. (1990). Item response theory, item calibration, and proficiency estimation. *Computerized adaptive testing: A primer*, 65-102.

Wallace, C. J. (1979). The effects of delayed rewards, social pressure, and frustration on the responses of opiate addicts. *NIDA Monograph Series*, 25, 6-25.

Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, 66(2), 98-112.

Wang, N., Johnson, W.L., Rizzo, P., Shaw, E., & Mayer, R.E. (2005). Experimental evaluation of polite interaction tactics for pedagogical agents. *Proceedings of the 10th international conference on Intelligent user interfaces* (12-19).

Waterhouse, I. K., & Child, I. L. (1953). Frustration and the quality of performance. *Journal of personality*, 21(3), 298-311.

Watson, B., Walker, N., Woytiuk, P., & Ribarsky, W. (2003, March). Maintaining usability during 3D placement despite delay. In *Virtual Reality, 2003. Proceedings. IEEE* (pp. 133-140). IEEE.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4), 548.

Wenger, E. (1987). Artificial Intelligence and Tutoring Systems. *Los Altos, CA: Morgan Kaufmann*

Wiethoff, M. A. A. G., Arnold, A. G., & Houwing, E. M. (1991). The value of psychophysiological measures in human-computer interaction. *Human-Aspects in*

*Computing: Design and Use of Interactive Systems and Work with Terminals.* Amsterdam: Elsevier, 661-665.

Wiggins, G. (2012). Seven keys to effective feedback. *Feedback for Learning*, 70(1), 10-16.

Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education*, 33(2), 153-169.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., & Picard, R. (2007). Emotional intelligence for computer tutors. *Workshop on Modeling and Scaffolding Affective Experiences to Impact Learning at 13th International Conference on Artificial Intelligence in Education*, Los Angeles, California.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognizing and responding to student affect. *International Journal of Learning Technology*, 4(3-4), 129-164.

Wu, P., Miller, C. A., Funk, H., & Vikili, V. (2010). Computational models of etiquette and Culture. Human-Computer Etiquette: Cultural Expectations and the Design Implications They Place on *Computers and Technology*, 63.

Yang, E., & Dorneich, M. C. (2015). The Effect of Time Delay on Emotion, Arousal, and Satisfaction in Human-Robot Interaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 443-447). SAGE Publications.

Yang, E. & Dorneich, M. C. (2016) Evaluation of Etiquette Strategies to Adapt Feedback in Affect-Aware Tutoring. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. In press.

Yang, E. & Dorneich, M. C. (in review) Evaluating Human-Automation Etiquette Strategies to Enhance Learning. *Journal of Educational Technology & Society*.

Zakharov, K., Mitrovic, A., & Ohlsson, S. (2005, May). Feedback micro-engineering in EER-Tutor. In *Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (pp. 718-725). IOS Press.

Zaman, B., & Shrimpton-Smith, T. (2006). The FaceReader: Measuring instant fun of use. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles* (pp. 457-460). ACM.

Zhong, P. (2013). Perception in Remote Navigation. Ph.D. Dissertation, *Iowa State University*.

Zoghbi S, Croft E, Kulic D, Van der Loos M (2009) Evaluation of affective state estimations using an on-line reporting device during human-robot interactions. In: *Proceedings of the 2009 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp 3742–3749.

APPENDIX A. [STUDY 1 EXPERIMENT MATERIALS (IRB #14-216)]

IRB #14-216 Approval

# IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

**Date:** 5/6/2014

**To:** Euijung Yang
3004 Black Engineering

**CC:** Dr. Michael Dorneich
3018 Black Engineering Bldg

**From:** Office for Responsible Research

**Title:** The Effect of Frustration on Cognitive Workload

**IRB ID:** 14-216

| | | | |
|---|---|---|---|
| **Approval Date:** | 5/6/2014 | **Date for Continuing Review:** | 5/4/2016 |
| **Submission Type:** | New | **Review Type:** | Expedited |

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**

- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required.

- **Obtain IRB approval prior to implementing <u>any</u> changes** to the study by submitting a Modification Form for Non-Exempt Research or Amendment for Personnel Changes form, as necessary.

- **Immediately inform the IRB of (1) all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.

- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.

- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Please be aware that IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.

Pre-experiment Survey

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

1. Gender
   - Male
   - Female

2. Age_____

3. Major_____

4. Do you have normal eye-sight (20/20) or corrected eyesight (with glasses or contact lenses)?
   - Yes
   - No

5. Are you using pace maker?
   - Yes
   - No

6. Which is your dominant hand?
   - Right hand
   - Left hand
   - Both hands

7. How long have you been playing video games?
   - Never
   - 6 months
   - 1 year
   - 2-5 years
   - 5-10 years
   - 10 or more years

8. How often (approximately) do you currently play video games?
   - never
   - daily
   - weekly
   - once a month
   - once in 6 months

- once a year
- less than once a year or never

9. What is your favorite genre of video game?
   - Action
   - Shooting
   - Role-playing
   - Flight
   - Racing
   - Sports
   - Military
   - Arcade
   - Maze
   - Puzzle
   - Other

10. How would you describe your current skill level (approximate) for playing video game?
    - Novice - video game is new to me
    - Beginner - I have played a few games but I am still learning
    - Intermediate - I have played enough games to know different strategies but I am still developing my skills
    - Expert - I have played enough games to know what strategies work best for me and when to implement those strategies

Post-trial Survey

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

1. How was the speed of the system while you operated robot?
    - Very Slow
    - Slow
    - Neutral
    - Fast
    - Very Fast


2. How was the smoothness of the system while you operated robot?
    - Very Rough
    - Rough
    - Neutral
    - Smooth
    - Very Smooth


3. Was feedback from system appropriate?
    - Very Inappropriate
    - Inappropriate
    - Neutral
    - Appropriate
    - Very Appropriate


4. How would you rate robotic control task?
    - Very Difficult
    - Difficult
    - Neutral
    - Easy
    - Very Easy


5. How much satisfaction did you experience as a result of this remote robot navigation?
    - Very Dissatisfied
    - Dissatisfied
    - Neutral
    - Satisfied
    - Very Satisfied

6. How did feel during operating robot?

| | Never | Rarely | Sometimes | Often | All of the Time |
|---|---|---|---|---|---|
| Anger | | | | | |
| Happiness | | | | | |
| Surprise | | | | | |
| Disgust | | | | | |
| Sadness | | | | | |
| Fear | | | | | |
| Frustration | | | | | |

7. Please describe your reaction to this experience.

Post-experiment Survey

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

1. What, if any, was the most difficult part of the robotic control task?

2. What, if any, was the easiest part of the robotic control task?

3. What kind of strategies did you use when navigating the robot?

4. Please describe 3 things most frustrating part during the robotic control task.

5. Please describe 3 things least frustrating part during the robotic control task.

## APPENDIX B. [STUDY 2 EXPERIMENT MATERIALS (IRB #15-142)]

IRB #15-142 Approval

# IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames. Iowa 50011-2207
515 294-4566
FAX 515 294-4267

**Date:** 4/13/2015

**To:** Euijung Yang  
3004 Black Engineering

**CC:** Dr. Michael Dorneich  
3018 Black Engineering Bldg

**From:** Office for Responsible Research

**Title:** Automation etiquette to mitigate negative emotions of students in learning environment

**IRB ID:** 15-142

| | | | |
|---|---|---|---|
| **Approval Date:** | 4/10/2015 | **Date for Continuing Review:** | 4/9/2017 |
| **Submission Type:** | New | **Review Type:** | Expedited |

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**

- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required.

- **Obtain IRB approval prior to implementing <u>any</u> changes** to the study by submitting a Modification Form for Non-Exempt Research or Amendment for Personnel Changes form, as necessary.

- **Immediately inform the IRB of (1) all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.

- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.

- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Please be aware that IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.

Pre-experiment Survey

Basic Demographics

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

1. Gender
   ☐ Female
   ☐ Male

2. Age_____

3. What is the highest grade or year of school you completed?
   ☐ Grade 12 or GED (High school graduate)
   ☐ College 1 year to 3 years (Some college of technical school) – Major:_____
   ☐ College 4 years (College graduate) – Major:_____
   ☐ Graduate School (Advance Degree) – Major:_____

4. How do you describe yourself? (please check the one option that best describes you)
   ☐ American Indian or Alaska Native
   ☐ Hawaiian or Other Pacific Islander
   ☐ Asian or Asian American
   ☐ Black or African American
   ☐ Hispanic or Latino
   ☐ Non-Hispanic White or Caucasian

5. Are you an international student?
   ☐ Domestic
   ☐ International – Please specify where you came from:_____

6. Are you using pace maker?   Y  /  N

7. Do you have normal eye-sight (20/20) or corrected eyesight (with glasses or contact lenses)?

   Y  /  N

8. Do you have a history of seizures?    Y  /  N

9. Approximately, how many hours to use computer a day? _____


Personality Test

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

10. Please check the options that you think the descriptions of yourself. (Jung, Myers, & Briggs, 1971)

| 10.1. How do you get your energy? | |
| --- | --- |
| ☐  are generally sociable | ☐  are generally quiet |
| ☐  are focused on the outer world | ☐  are focused on their inner world |
| ☐  get energy by spending time with others | ☐  get energy by spending time alone |
| ☐  talk a lot & start conversations | ☐  mostly listen & wait for others to talk first |
| ☐  speak first, then think | ☐  think first, then speak |
| ☐  are quick to take action | ☐  are slow to take action |
| ☐  have many friends & many interests | ☐  have a few deep friendships & refined interests |

| 10.2. How do you see the world & gather information? | |
| --- | --- |
| ☐  have finely-tuned five senses | ☐  use their "sixth sense" |
| ☐  pay attention to the details | ☐  see the "big picture" |
| ☐  focus on what is real (in the present) | ☐  focus on what is possible (in the future) |
| ☐  think in concrete terms | ☐  think in abstract terms |
| ☐  like practical things | ☐  like theories |
| ☐  like to do (make) | ☐  like to dream (design) |
| ☐  are accurate and observant | ☐  are creative and imaginative |
| ☐  prefer to do things the established way | ☐  prefer to try out new ideas |

| 10.3. How do you make your decisions? | |
| --- | --- |
| ☐  mostly use their head | ☐  mostly use their heart |
| ☐  make decisions based on logic | ☐  make decisions based on their values |
| ☐  are more interested in things & ideas | ☐  are more interested in people & emotions |
| ☐  treat everybody the same (emphasizing fairness) | ☐  treat people according to their situation (emphasizing compassion) |
| ☐  are more scientific in describing the world | ☐  are more poetic in describing the world |

| 10.4. How much do you like to plan ahead? | |
| --- | --- |
| ☐  are organized and structured | ☐  are casual and relaxed |
| ☐  make plans in advance | ☐  prefer to "go with the flow" |
| ☐  keep to the plan | ☐  are able to change and adapt quickly |
| ☐  like to be in control of their life | ☐  like to simply let life happen |
| ☐  want to finalize decisions | ☐  want to find more information |

11. Please mark your preference of conversation style.

**Directness**

Indirect                                                                 Direct

**Speaking pace**

Slow                                                                       Fast

**Strong voice**

Soft                                                                      Strong

**Confidence**

Low                                                                        High

**Briefness**

Brief                                                                     Lengthy

**Personal conversation**

Non personal                                                            Personal

**Energetic conversation**

Lethargic                                                              Energetic

**Friendliness**

Aloof                                                                                                    Friendly

**Loudness**

| | | | | | | | | | | | | | | | | | | | | | | | |

Calm                                                                                                        Loud

**Eye contact**

| | | | | | | | | | | | | | | | | | | | | | | | |

Never                                                                                                 All the time

Interaction Style Preference

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

12. There are definitions and short example sentences of four different etiquette strategies in interaction. Please mark your preference. (Brown & Levinson, 1987)

- Strategy Name: **Bald**
- Definition: Bald strategy is a direct way of saying things, without any minimization to the imposition, in a direct, clear, unambiguous and concise way, for example "Do X!".

**Preference**

| | | | | | | | | | | | | | | | | | | | | | | | |

Dislike                                                                                                      Like

- Strategy Name: **Positive politeness**
- Definition: Positive politeness is to minimize the social distance between speaker and hearer by expressing statements of friendship, solidarity, and compliments.

**Preference**

| | | | | | | | | | | | | | | | | | | | | | | | |

Dislike                                                                                                      Like

- Strategy Name: **Negative politeness**

- Definition: Negative politeness is to be respectful however, speaker also assumes that s/he is in some way imposing on hearer. Examples would be to say, "I don't want to bother you but." or "I was wondering if."

**Preference**

Dislike                                                 Like

- Strategy Name: **Off-record**
- Definition: Off record utterances are essential in indirect use of language. One says something that is rather general. In this case, the hearer must make some inference to recover what was intended.

**Preference**

Dislike                                                 Like

13. What would you do if you see a cup of pens on your teacher's desk, and you want to use one, would you:
    - ☐ Say to teacher, "I want to use one of those!"
    - ☐ Say to teacher, "Is it okay, if I use one of those pens?"
    - ☐ Say to teacher, "I'm sorry to bother you but, I just wanted to ask you if I could use one of those pens?"
    - ☐ Say to myself, "Hmm, I sure could use a blue pen right now."

14. What would you do if you want to read a book in quiet living room, but your roommate is listening music by using speaker, would you:
    - ☐ Say to roommate, "I want to read a book in here."
    - ☐ Say to roommate, "Why don't you use the headphone?"
    - ☐ Say to roommate, "I'm sorry to bother you but, I just wanted to read a book in here. Could you please use the headphone instead of the speaker?"
    - ☐ Say to myself, "Hmm, I think the music is a little bit loud."

15. What would you do if your younger sister or brother is struggling with a simple math problem that you know how to solve, and you want to help her or him, would you:
    - ☐ Say to sister, "I want to help you out how to solve it."
    - ☐ Say to sister, "Why don't you try to solve this problem with me? Let's do it together."
    - ☐ Say to sister, "I'm sorry to bother you but, I just wanted to help you out."
    - ☐ Say to myself, "Hmm, I learned how to solve that problem."

16. What would you do if you are lost directions while you are traveling, and you want to ask right direction to people to go to city hall, would you:
    ☐ Say to stranger, "Hi, I want to know how to go to city hall!"
    ☐ Say to stranger, "Hi, do you know how to go to city hall?"
    ☐ Say to stranger, "Excuse me, I'm sorry to bother you, could you give me directions to go to city hall?"
    ☐ Say to stranger, "Hi, I think I am lost directions to go to city hall."

17. What would you do if the conference room is too cold when you are with your business partner who is sitting near the air heater button, and you want to ask to turn it on, would you:
    ☐ Say to business partner, "I want you to turn on the air heater."
    ☐ Say to business partner, "If it is okay, why don't you turn on the air heater?"
    ☐ Say to business partner, "I'm sorry to bother you but, I just wanted to ask you if you can turn on the air heater."
    ☐ Say to myself, "Hmm, I think it's too cold now."

18. What would you do if your best friend is wearing his or her shirt inside out, and you want to let him or her know, would you:
    ☐ Say to friend, "You are wearing your shirt inside out."
    ☐ Say to friend, "Why don't you check your shirt? It is inside out now."
    ☐ Say to friend, "I'm sorry to say this but, I just wanted to let you know your shirt is inside out."
    ☐ Say to friend, "Hmm, I think something is wrong with your shirt."

19. What would you do if you have physics homework which are pretty difficult, and you want to ask how to solve to your class mate, would you:
    ☐ Say to class mate, "I want you to help me to solve this."
    ☐ Say to class mate, "Do you know how to solve this problems? Let's solve it together."
    ☐ Say to class mate, "I'm sorry to bother you but, I just wanted to ask how to solve this physics problems. Could you please let me know if you know it?"
    ☐ Say to myself, "Hmm, I think the physics problems are too difficult."

20. What would you do if you work on school project with your team members, and are wondering how the other team members are doing, would you:
    ☐ Say to other team member, "How is your team work going?"
    ☐ Say to other team member, "If it is okay, do you want to let me know how your team work is going?"
    ☐ Say to other team member, "I'm sorry to bother you but, I was just wondering how your team work is going. Could please you let me know about it?"
    ☐ Say to other team member, "Hmm, I think your team is doing well."

21. What would you do if you are already done with your portion of team project while other team members are not yet, and you want to check how they are doing, would you:
- ☐ Say to team member, "I wonder how your part is going."
- ☐ Say to team member, "Do you want to check together how we are doing with our team work?"
- ☐ Say to team member, "I'm sorry to bother you but, I just wanted to know how your part is going. Could you please let me know about it?"
- ☐ Say to myself, "Hmm, I hope our project is going well."

Learning Environments

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

22. Please mark your preference of teaching style in a learning environment.

**Authority (lecture style)**

Dislike                                                                 Like

**Demonstrator (coach style)**

Dislike                                                                 Like

**Facilitator (activity style)**

Dislike                                                                 Like

**Delegator (group style)**

Dislike                                                                 Like

**Hybrid (blended style)**

Dislike                                                                 Like

23. As a student, please check your preference of the teacher's teaching style in learning environment. (Keeley, Smith, & Buskist, 2006)

**Humble**

Dislike                                                                                     Like

**Sensitive and persistent**

Dislike                                                                                     Like

**Strives to be a better teacher**

Dislike                                                                                     Like

**Respectful**

Dislike                                                                                     Like

**Encourages and cares for students**

Dislike                                                                                     Like

**Enthusiastic about teaching**

Dislike                                                                                     Like

**Good listener**

Dislike                                                                                     Like

**Flexible/open-minded**

Dislike                                                                                          Like

**Understanding**

Dislike                                                                                          Like

**Happy/positive attitude/humorous**

Dislike                                                                                          Like

**Approachable/personable**

Dislike                                                                                          Like

**Rapport**

Dislike                                                                                          Like

**Provides constructive feedback**

Dislike                                                                                          Like

**Realistic expectations of students**

Dislike                                                                                          Like

**Fair testing and grading**

Dislike                                                                                          Like

**Creative and interesting**

Dislike                                                                                              Like

**Effective communicator**

Dislike                                                                                              Like

**Professional**

Dislike                                                                                              Like

**Knowledgeable about subject matter**

Dislike                                                                                              Like

**Prepared**

Dislike                                                                                              Like

**Punctuality/manages class time**

Dislike                                                                                              Like

**Establishes academic term goals**

Dislike                                                                                              Like

**Promotes intellectually stimulating**

Dislike                                                                                              Like

**Confident**

Dislike                                                                                                                                    Like

**Presents current information**

Dislike                                                                                                                                    Like

**Accessible**

Dislike                                                                                                                                    Like

**Authoritative**

Dislike                                                                                                                                    Like

**Technologically competent**

Dislike                                                                                                                                    Like

- If you have any other expectations not stated above, please describe it.

24. Please check your familiarity of online learning environment.

**Online classes**

Low                                                                                                                                       High

**E-learning system**

Low                                                                                              High

**Intelligent tutoring system**

Low                                                                                              High

**Tutorial video**

Low                                                                                              High

- If you have any other experiences not stated above, please describe it.

Task

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

25. How would you describe your current skill level of math?

**Algebra**

Poor                                                                                            Good

**Linear Algebra**

Poor                                                                                            Good

**Geometry**

Poor                                                                 Good

**Trigonometry**

Poor                                                                 Good

**Pre-calculus**

Poor                                                                 Good

**Calculus**

Poor                                                                 Good

**Statistics**

Poor                                                                 Good

**Probability**

Poor                                                                 Good

26. When was the last time did you take math class?
    ☐ Within 1 year
    ☐ Within 2 year
    ☐ Within 3 year
    ☐ Within 4 year
    ☐ Within 5 year

Emotional States

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

27. How do you describe your current emotional status?

**Happy**

Disagree                                                                                          Agree

**Bored**

Disagree                                                                                          Agree

**Sad**

Disagree                                                                                          Agree

**Stressed**

Disagree                                                                                          Agree

**Frustrated**

Disagree                                                                                          Agree

**Angry**

Disagree                                                                                          Agree

Post-trial Survey

Feedback Effectiveness

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

1. Was feedback from system appropriate?

**Appropriateness**

Disagree                                                                                                      Agree

2. Was feedback from system effective to solve the task?

**Effectiveness**

Disagree                                                                                                      Agree

3. How motivated do you feel to continue working on tasks?

**Motivation**

Disagree                                                                                                      Agree

4. How much satisfaction did you experience based on system's feedback?

**Satisfaction on system's feedback**

Disagree                                                                                                      Agree

Task Performance

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

5. How confident do you feel about your performance during the task?

**Confidence**

Disagree                                                                                                    Agree

6. How successful do you feel about your performance of the task?

**Success**

Disagree                                                                                                    Agree

7. Was you satisfied with your performance of the task?

**Satisfaction on task performance**

Disagree                                                                                                    Agree

Emotional States

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

8. How do you describe your emotional status after this trial?

**Happy**

Disagree                                                                                                    Agree

**Bored**

Disagree                                                                                                    Agree

**Sad**

Disagree                                                                                                    Agree

**Stressed**
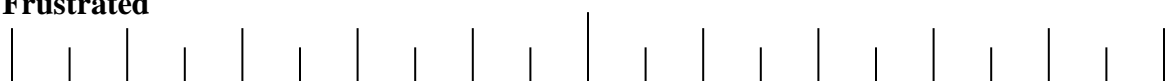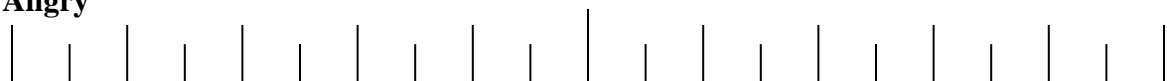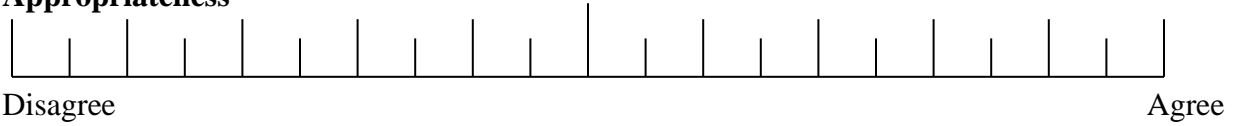
Disagree                                                                                                    Agree

**Frustrated**

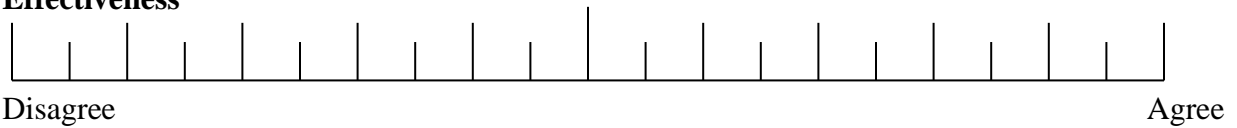Disagree                                                                                                    Agree

**Angry**

Disagree                                                                                                    Agree

Post-experiment Survey

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

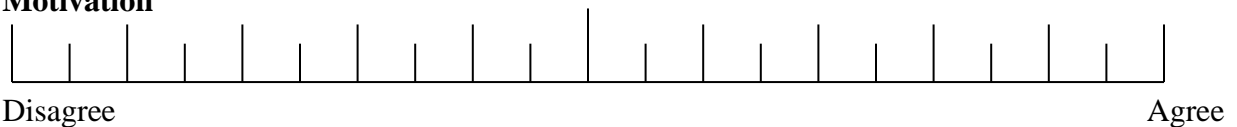1. What, if any, was the most difficult part of the task?

   1.1.Why do you think that is the most difficult part?

2. What, if any, was the easiest part of the task?

   2.1.Why do you think that is the easiest part?

3.  What kind of strategies did you use when solving the task?

```



```

4.  Please describe 3 things most frustrating part during the task.

```



```

5.  Please describe 3 things least frustrating part during the task.

```



```

## APPENDIX C. [STUDY 3 EXPERIMENT MATERIALS (IRB #16-004)]

IRB #16-004 Approval

# IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

Date: 2/19/2016

To: Euijung Yang
3004 Black Engineering

CC: Dr. Michael Dorneich
3018 Black Engineering Bldg

From: Office for Responsible Research

Title: Evaluation of Adaptive Tutoring System

IRB ID: 16-004

| | | | |
|---|---|---|---|
| Approval Date: | 2/19/2016 | Date for Continuing Review: | 2/18/2018 |
| Submission Type: | New | Review Type: | Expedited |

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- Use only the approved study materials in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**

- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required.

- **Obtain IRB approval prior to implementing <u>any</u> changes** to the study by submitting a Modification Form for Non-Exempt Research or Amendment for Personnel Changes form, as necessary.

- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.

- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.

- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Please be aware that IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.

Pre-experiment Survey

Basic Demographics

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

1. Gender
   ☐ Female
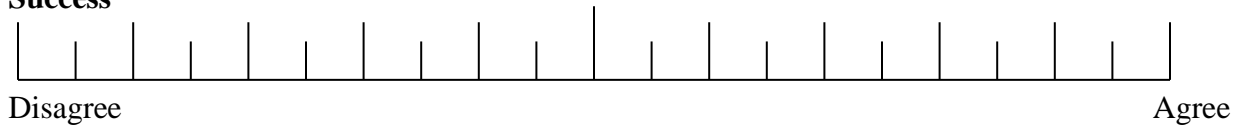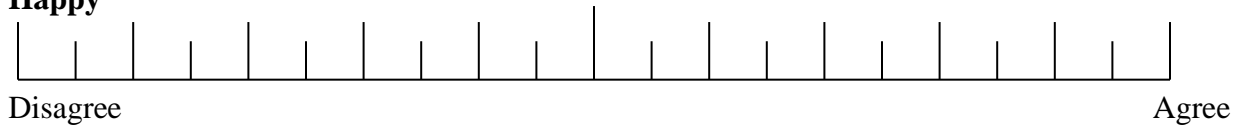   ☐ Male

2. Age_____

3. What is the highest grade or year of school you completed?
   ☐ Grade 12 or GED (High school graduate)
   ☐ College 1 year to 3 years (Some college of technical school) – Major:_____
   ☐ College 4 years (College graduate) – Major:_____
   ☐ Graduate School (Advance Degree) – Major:_____

4. How do you describe yourself? (please check the one option that best describes you)
   ☐ American Indian or Alaska Native
   ☐ Hawaiian or Other Pacific Islander
   ☐ Asian or Asian American
   ☐ Black or African American
   ☐ Hispanic or Latino
   ☐ Non-Hispanic White or Caucasian

5. Are you an international student?
   ☐ Domestic
   ☐ International – Please specify where you came from:_____

6. Are you using pace maker?   Y  /  N

7. Do you have normal eye-sight (20/20) or corrected eyesight (with glasses or contact lenses)?

   Y  /  N

8. Do you have a history of seizures?    Y  /  N

9. Approximately, how many hours to use computer a day? _____


Personality Test

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

    10. Please check the options that you think the descriptions of yourself. (Jung, Myers, & Briggs, 1971)

| 10.1. How do you get your energy? | |
|---|---|
| ☐ are generally sociable | ☐ are generally quiet |
| ☐ are focused on the outer world | ☐ are focused on their inner world |
| ☐ get energy by spending time with others | ☐ get energy by spending time alone |
| ☐ talk a lot & start conversations | ☐ mostly listen & wait for others to talk first |
| ☐ speak first, then think | ☐ think first, then speak |
| ☐ are quick to take action | ☐ are slow to take action |
| ☐ have many friends & many interests | ☐ have a few deep friendships & refined interests |

| 10.2. How do you see the world & gather information? | |
|---|---|
| ☐ have finely-tuned five senses | ☐ use their "sixth sense" |
| ☐ pay attention to the details | ☐ see the "big picture" |
| ☐ focus on what is real (in the present) | ☐ focus on what is possible (in the future) |
| ☐ think in concrete terms | ☐ think in abstract terms |
| ☐ like practical things | ☐ like theories |
| ☐ like to do (make) | ☐ like to dream (design) |
| ☐ are accurate and observant | ☐ are creative and imaginative |
| ☐ prefer to do things the established way | ☐ prefer to try out new ideas |

| 10.3. How do you make your decisions? | |
|---|---|
| ☐ mostly use their head | ☐ mostly use their heart |
| ☐ make decisions based on logic | ☐ make decisions based on their values |
| ☐ are more interested in things & ideas | ☐ are more interested in people & emotions |
| ☐ treat everybody the same (emphasizing fairness) | ☐ treat people according to their situation (emphasizing compassion) |
| ☐ are more scientific in describing the world | ☐ are more poetic in describing the world |

| 10.4. How much do you like to plan ahead? | |
|---|---|
| ☐ are organized and structured | ☐ are casual and relaxed |
| ☐ make plans in advance | ☐ prefer to "go with the flow" |
| ☐ keep to the plan | ☐ are able to change and adapt quickly |
| ☐ like to be in control of their life | ☐ like to simply let life happen |
| ☐ want to finalize decisions | ☐ want to find more information |

11. Please mark your preference of conversation style.

**Directness**

Indirect                                                                                                    Direct

**Speaking pace**

Slow                                                                                                            Fast

**Strong voice**

Soft                                                                                                          Strong

**Confidence**

Low                                                                                                            High

**Briefness**

Brief                                                                                                       Lengthy

**Personal conversation**

Non personal                                                                                        Personal

**Energetic conversation**

Lethargic                                                                                           Energetic

**Friendliness**

Aloof                                                                                           Friendly

**Loudness**

| | | | | | | | | | | | | | | | | | | | | | |

Calm                                                                                              Loud

**Eye contact**

| | | | | | | | | | | | | | | | | | | | | | |

Never                                                                                    All the time

Interaction Style Preference

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

12. There are definitions and short example sentences of four different etiquette strategies in interaction. Please mark your preference. (Brown & Levinson, 1987)

- Strategy Name: **Bald**
- Definition: Bald strategy is a direct way of saying things, without any minimization to the imposition, in a direct, clear, unambiguous and concise way, for example "Do X!".

**Preference**

| | | | | | | | | | | | | | | | | | | | | | |

Dislike                                                                                            Like

- Strategy Name: **Positive politeness**
- Definition: Positive politeness is to minimize the social distance between speaker and hearer by expressing statements of friendship, solidarity, and compliments.

**Preference**

| | | | | | | | | | | | | | | | | | | | | | |

Dislike                                                                                            Like

- Strategy Name: **Negative politeness**

- Definition: Negative politeness is to be respectful however, speaker also assumes that s/he is in some way imposing on hearer. Examples would be to say, "I don't want to bother you but." or "I was wondering if."

**Preference**

| | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

Dislike                                                                                                    Like

- Strategy Name: **Off-record**
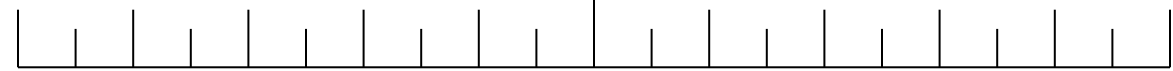- Definition: Off record utterances are essential in indirect use of language. One says something that is rather general. In this case, the hearer must make some inference to recover what was intended.

**Preference**

| | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

Dislike                                                                                                    Like

13. What would you do if you see a cup of pens on your teacher's desk, and you want to use one, would you:
    ☐ Say to teacher, "I want to use one of those!"
    ☐ Say to teacher, "Is it okay, if I use one of those pens?"
    ☐ Say to teacher, "I'm sorry to bother you but, I just wanted to ask you if I could use one of those pens?"
    ☐ Say to myself, "Hmm, I sure could use a blue pen right now."

14. What would you do if you want to read a book in quiet living room, but your roommate is listening music by using speaker, would you:
    ☐ Say to roommate, "I want to read a book in here."
    ☐ Say to roommate, "Why don't you use the headphone?"
    ☐ Say to roommate, "I'm sorry to bother you but, I just wanted to read a book in here. Could you please use the headphone instead of the speaker?"
    ☐ Say to myself, "Hmm, I think the music is a little bit loud."

15. What would you do if your younger sister or brother is struggling with a simple math problem that you know how to solve, and you want to help her or him, would you:
    ☐ Say to sister, "I want to help you out how to solve it."
    ☐ Say to sister, "Why don't you try to solve this problem with me? Let's do it together."
    ☐ Say to sister, "I'm sorry to bother you but, I just wanted to help you out."
    ☐ Say to myself, "Hmm, I learned how to solve that problem."

16. What would you do if you are lost directions while you are traveling, and you want to ask right direction to people to go to city hall, would you:
☐ Say to stranger, "Hi, I want to know how to go to city hall!"
☐ Say to stranger, "Hi, do you know how to go to city hall?"
☐ Say to stranger, "Excuse me, I'm sorry to bother you, could you give me directions to go to city hall?"
☐ Say to stranger, "Hi, I think I am lost directions to go to city hall."

17. What would you do if the conference room is too cold when you are with your business partner who is sitting near the air heater button, and you want to ask to turn it on, would you:
☐ Say to business partner, "I want you to turn on the air heater."
☐ Say to business partner, "If it is okay, why don't you turn on the air heater?"
☐ Say to business partner, "I'm sorry to bother you but, I just wanted to ask you if you can turn on the air heater."
☐ Say to myself, "Hmm, I think it's too cold now."

18. What would you do if your best friend is wearing his or her shirt inside out, and you want to let him or her know, would you:
☐ Say to friend, "You are wearing your shirt inside out."
☐ Say to friend, "Why don't you check your shirt? It is inside out now."
☐ Say to friend, "I'm sorry to say this but, I just wanted to let you know your shirt is inside out."
☐ Say to friend, "Hmm, I think something is wrong with your shirt."

19. What would you do if you have physics homework which are pretty difficult, and you want to ask how to solve to your class mate, would you:
☐ Say to class mate, "I want you to help me to solve this."
☐ Say to class mate, "Do you know how to solve this problems? Let's solve it together."
☐ Say to class mate, "I'm sorry to bother you but, I just wanted to ask how to solve this physics problems. Could you please let me know if you know it?"
☐ Say to myself, "Hmm, I think the physics problems are too difficult."

20. What would you do if you work on school project with your team members, and are wondering how the other team members are doing, would you:
☐ Say to other team member, "How is your team work going?"
☐ Say to other team member, "If it is okay, do you want to let me know how your team work is going?"
☐ Say to other team member, "I'm sorry to bother you but, I was just wondering how your team work is going. Could please you let me know about it?"
☐ Say to other team member, "Hmm, I think your team is doing well."

21. What would you do if you are already done with your portion of team project while other team members are not yet, and you want to check how they are doing, would you:
   ☐ Say to team member, "I wonder how your part is going."
   ☐ Say to team member, "Do you want to check together how we are doing with our team work?"
   ☐ Say to team member, "I'm sorry to bother you but, I just wanted to know how your part is going. Could you please let me know about it?"
   ☐ Say to myself, "Hmm, I hope our project is going well."

Learning Environments

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.
   22. Please mark your preference of teaching style in a learning environment.

**Authority (lecture style)**

Dislike                                                         Like

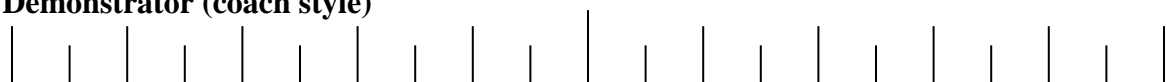**Demonstrator (coach style)**

Dislike                                                         Like
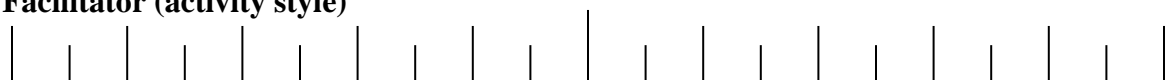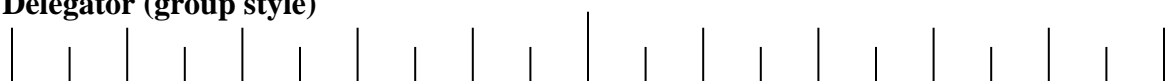
**Facilitator (activity style)**

Dislike                                                         Like

**Delegator (group style)**

Dislike                                                         Like

**Hybrid (blended style)**

Dislike                                                         Like

23. As a student, please check your preference of the teacher's teaching style in learning
environment. (Keeley, Smith, & Buskist, 2006)

**Humble**

Dislike                                                                                    Like

**Sensitive and persistent**

Dislike                                                                                    Like

**Strives to be a better teacher**

Dislike                                                                                    Like

**Respectful**

Dislike                                                                                    Like

**Encourages and cares for students**

Dislike                                                                                    Like

**Enthusiastic about teaching**

Dislike                                                                                    Like

**Good listener**

Dislike                                                                                    Like

**Flexible/open-minded**

Dislike                                                                                    Like

**Understanding**

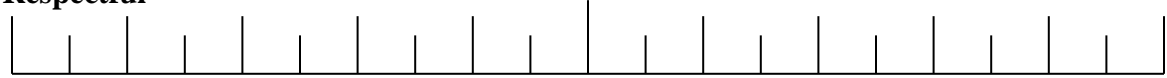Dislike                                                                                    Like

**Happy/positive attitude/humorous**

Dislike                                                                                    Like

**Approachable/personable**

Dislike                                                                                    Like

**Rapport**

Dislike                                                                                    Like

**Provides constructive feedback**

Dislike                                                                                    Like

**Realistic expectations of students**

Dislike                                                                                    Like

**Fair testing and grading**

Dislike                                                                                    Like

**Creative and interesting**

Dislike                                                                Like

**Effective communicator**

Dislike                                                                Like

**Professional**

Dislike                                                                Like

**Knowledgeable about subject matter**
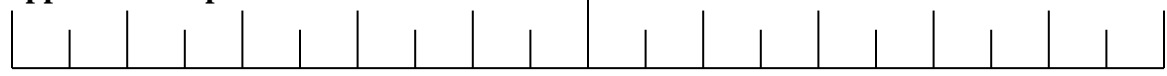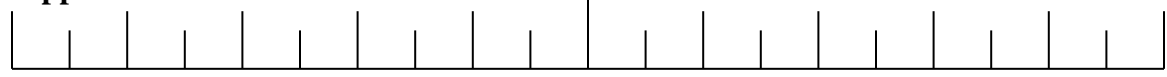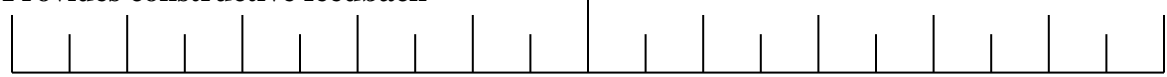
Dislike                                                                Like

**Prepared**

Dislike                                                                Like

**Punctuality/manages class time**

Dislike                                                                Like

**Establishes academic term goals**

Dislike                                                                Like

**Promotes intellectually stimulating**

Dislike                                                                Like

**Confident**

|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|

Dislike                                                                                                    Like

**Presents current information**

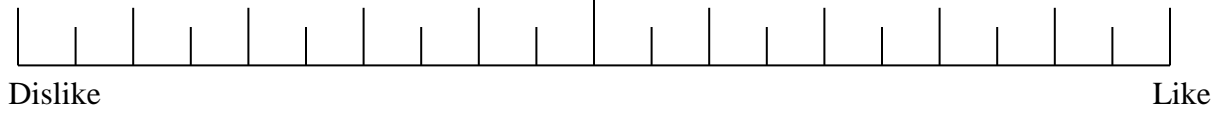|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|

Dislike                                                                                                    Like

**Accessible**

|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|

Dislike                                                                                                    Like

**Authoritative**

|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|

Dislike                                                                                                    Like

**Technologically competent**

|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|

Dislike                                                                                                    Like

- If you have any other expectations not stated above, please describe it.

```


```

24. Please check your familiarity of online learning environment.

**Online classes**

|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|

Low                                                                                                       High

**E-learning system**

Low                                                                                          High

**Intelligent tutoring system**

Low                                                                                          High

**Tutorial video**

Low                                                                                          High

- If you have any other experiences not stated above, please describe it.

Task

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

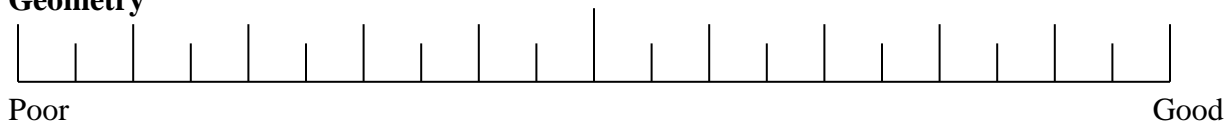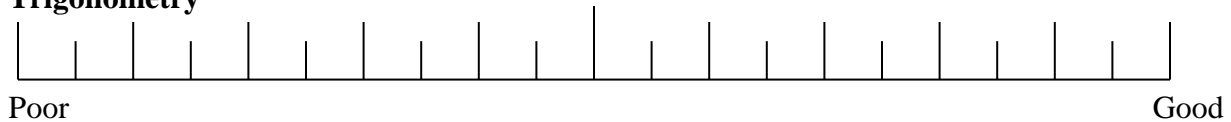25. How would you describe your current skill level of math?

**Algebra**
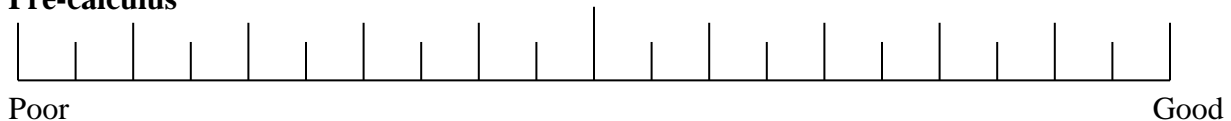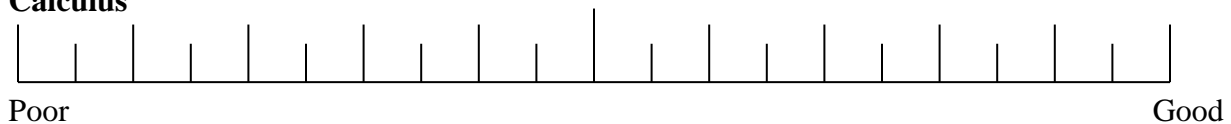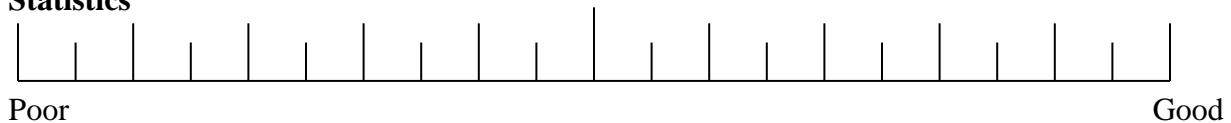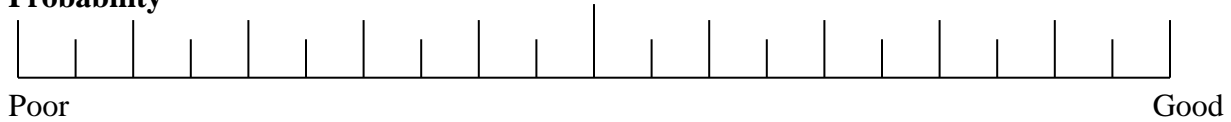
Poor                                                                                         Good

**Linear Algebra**

Poor                                                                                         Good

**Geometry**

| | | | | | | | | | | | | | | | | | | | | | | | | | |

Poor        Good

**Trigonometry**

| | | | | | | | | | | | | | | | | | | | | | | | | | |

Poor        Good

**Pre-calculus**

| | | | | | | | | | | | | | | | | | | | | | | | | | |

Poor        Good

**Calculus**

| | | | | | | | | | | | | | | | | | | | | | | | | | |

Poor        Good

**Statistics**

| | | | | | | | | | | | | | | | | | | | | | | | | | |

Poor        Good

**Probability**

| | | | | | | | | | | | | | | | | | | | | | | | | | |

Poor        Good

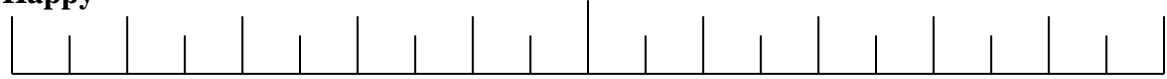26. When was the last time did you take math class?
   ☐ Within 1 year
   ☐ Within 2 year
   ☐ Within 3 year
   ☐ Within 4 year
   ☐ Within 5 year

Emotional States

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

27. How do you describe your current emotional status?

**Happy**

Disagree                                                                                          Agree

**Bored**

Disagree                                                                                          Agree

**Sad**

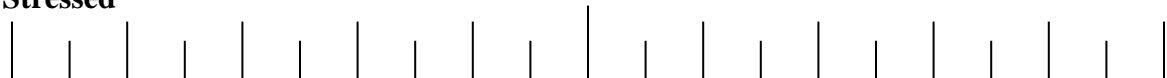Disagree                                                                                          Agree

**Stressed**

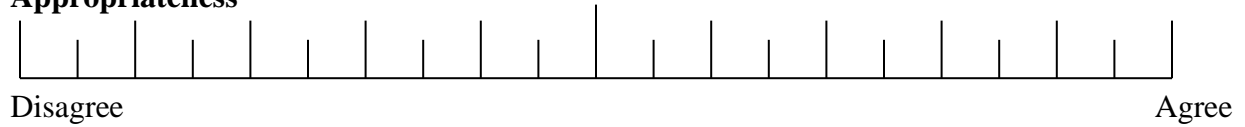Disagree                                                                                          Agree

**Frustrated**

Disagree                                                                                          Agree

**Angry**

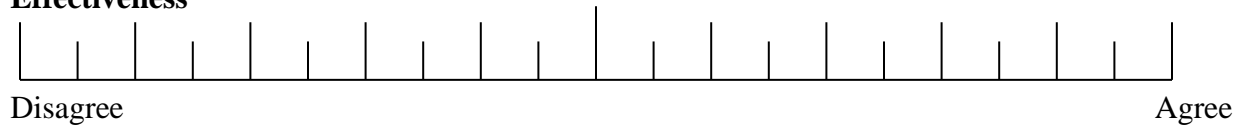Disagree                                                                                          Agree

Post-problem Survey

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.
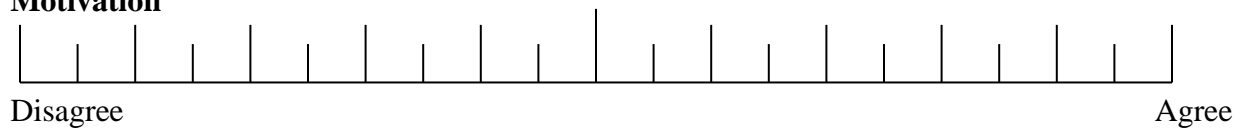
1. Was feedback from system appropriate?
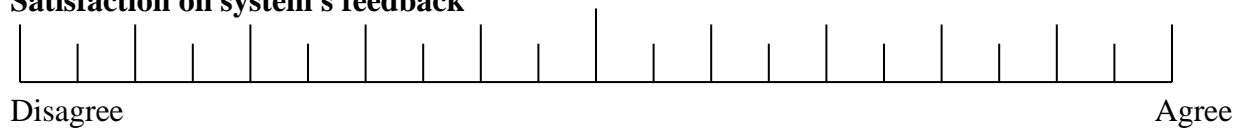
**Appropriateness**

Disagree                                                                                    Agree

2. Was feedback from system effective to solve the task?

**Effectiveness**

Disagree                                                                                    Agree

3. How motivated do you feel to continue working on tasks?

**Motivation**

Disagree                                                                                    Agree

4. How much satisfaction did you experience based on system's feedback?

**Satisfaction on system's feedback**

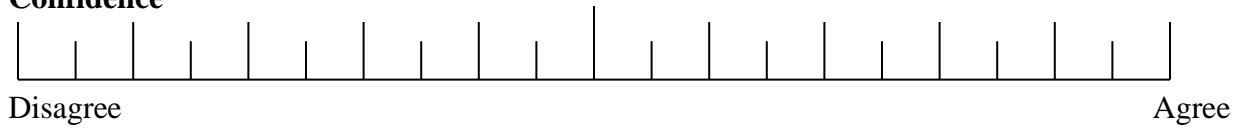Disagree                                                                                    Agree

Task Performance

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.
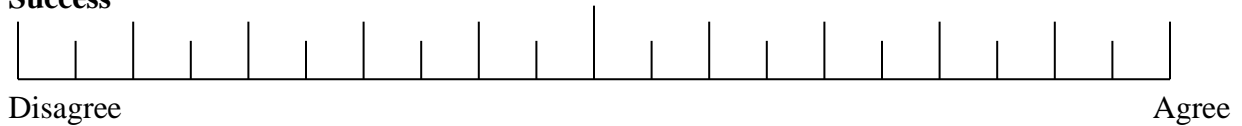
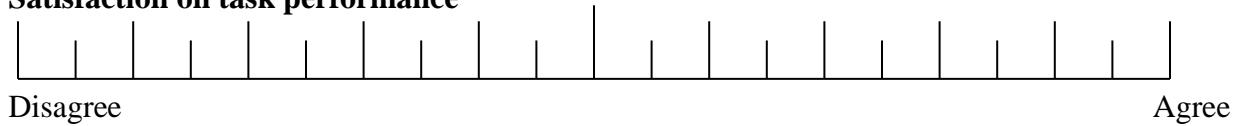5. How confident do you feel about your performance during the task?

**Confidence**

Disagree                                                                                                          Agree

6. How successful do you feel about your performance of the task?

**Success**

Disagree                                                                                                          Agree

7. Was you satisfied with your performance of the task?

**Satisfaction on task performance**

Disagree                                                                                                          Agree
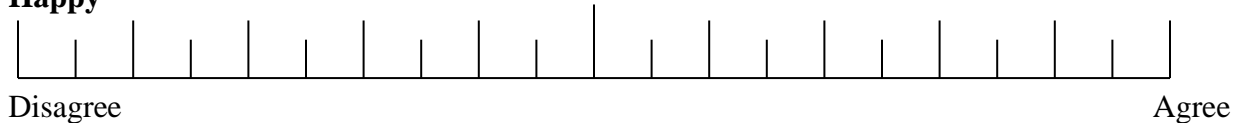
Emotional States

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

8. How do you describe your emotional status after this trial?

**Happy**

Disagree                                                                                                          Agree

**Bored**

Disagree                                                                                     Agree

**Sad**

Disagree                                                                                     Agree

**Stressed**
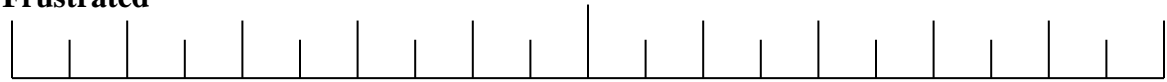
Disagree                                                                                     Agree

**Frustrated**

Disagree                                                                                     Agree

**Angry**

Disagree                                                                                     Agree

Post-experiment Survey

Please select the answer from the choices provided. All the information gathered in this study will be kept confidential.

1. What, if any, was the most difficult part of the task?

1.1 Why do you think that is the most difficult part?

2. What, if any, was the easiest part of the task?

2.1 Why do you think that is the easiest part?

3. What kind of strategies did you use when solving the task?

4. Please describe 3 things most frustrating part during the task.

5. Please describe 3 things least frustrating part during the task.